DESIGN OF HIGH-YIELDING NON-ENZYMATIC DNA AND RNA LIGATION SYSTEMS

A Dissertation
Presented to
The Academic Faculty

by

Chiamaka Obianyor

In Partial Fulfillment
of the Requirements for the Degree
Doctorate in Chemical Engineering in the
School of Chemical and Biomolecular Engineering

Georgia Institute of Technology
May 2021

COPYRIGHT © 2021 BY CHIAMAKA OBIANYOR

DESIGN OF HIGH-YIELDING NON-ENZYMATIC DNA AND RNA LIGATION SYSTEMS

Approved by:

Dr. Martha Grover, Advisor
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Mark Styczynski
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Nicholas Hud, co-Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Corey Wilson
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Anant Paravastu
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Date Approved: [March 24, 2021]

This Thesis is dedicated to my siblings, Chibuike Obianyor, Chiedozie Obianyor, and Chidiogo Obianyor. Dream no little dreams.

ACKNOWLEDGEMENTS

There are countless people who have labored tirelessly to get me to where I am today, and I will not be able to thank them all. So, I say thank you first and foremost to those whose names do not appear in this rather short piece of text. Your friendships, kindness, warmth, and love for me over the years mean a lot more than words can ever express and I will forever cherish the memories that we built together.

I would also like to thank my parents, Nneka Obianyor and Dr. Osita Obianyor, who have done an outstanding job of always making sure I lacked nothing, both in deed and in kind. Thank you to my dad for teaching me the value of hard work early, and to my mum for being not more than a phone call away when I needed her, which was nearly all the time. To my siblings, whom this Thesis is dedicated to, I would like to thank you all for being a ray of sunshine when the path seemed dark. Your laughter, dedication, and grit in all your endeavors pushed me to endure when times were hard. To my entire family, especially the large array of cousins, aunts, and uncles who provided a place for me to call home away from home, whose homes were filled with laughter and kindness, and who stepped in as the situation required, I am forever indebted to you all.

To my Ph.D. advisors, Martha Grover, and Nick Hud, thank you for always challenging me and accepting nothing but the best from me. Thank you for giving me a place and the space to grow as I matured as a researcher in a field that was at first entirely new to me. Thank you to my committee members, Anant Paravastu, Corey Wilson, and Mark Styczynski for helping expand my thinking and perspective over the years, your feedback was invaluable. I would also like to thank my undergraduate research advisor at

Texas Tech, Ted Wiesner, I learned to do research under your guidance and subsequently fell in love with the process of science and discovery. Thank you for teaching me in those early crucial years how to demand more from myself.

Finally, I would like to thank the countless lab members and collaborators within the Center for Chemical Evolution who went above and beyond for me during my graduate years. Christine Conwell, Shantel Floyd, all the Hud's buds I encountered; especially Adriana Lozoya-Colinas, David Fialho, Martin C, Moran Frenkel-Pinter, Suneesh Karunakaran, Sreejith Mangalath, Tyler Roche, Bryce E. Clifton, Bradley Burcar, Christine He, the honorary Hud's Bud Rio Febrian, all the Grover group members over the years, especially Stefani Kocevska, and last but certainly not least, our lab manager Gary Newnam. Gary, your role in my research endeavors in lab can definitely not be overstated. Thank you all for providing a nourishing, and welcoming environment for a young and curious researcher like me to thrive.

To all the black girls out there who are wondering whether there is a place for them in science and engineering. There is. You belong. Until next time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

NA	Nucleic Acids
bya	Billion years ago
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide or water soluble carbodiimide
T_m	Melting temperature
nt	Nucleotide
DAP	diamidophosphate

SUMMARY

The demonstration of enzyme-free replication of nucleic acids, in particular RNA, will lend credence to the RNA world hypothesis which states that RNA was the first polymer on the prebiotic earth. One major challenge that currently still exists is how substrates can undergo ligation (i.e. form a covalent bond) in the absence of enzymes. The use of activated species such as imidazole and water soluble carbodiimides have been proposed as means for ligation. However, obtaining the activating agents by a prebiotic route is an active area of research with no consensus yet on an exact pathway. Cyclic phosphate intermediates offer a potential route to prebiotic ligation of RNA substrates, but these reactions are often marred by low yields. In order to demonstrate true enzyme-free replication, a prebiotically plausible route to non-enzymatic ligation must be developed.

In this Thesis, a systematic non-enzymatic ligation method is described for both DNA and RNA ligating systems. DNA is used as a model system to investigate the limits of non-enzymatic ligation in a system that is prone to high yields, unlike the RNA cyclic phosphate system. This Thesis demonstrates how several factors such as binding of the substrates to the templates, side product reactions, and stacking of the bases at the ligation site can affect yields in DNA ligation. Applying these findings to the RNA cyclic phosphate ligation system, the role of sequences and base-pairing at the ligation nick, in addition to cyclic phosphate hydrolysis, were emphasized as potential barriers for attaining high yielding RNA systems in a prebiotic world.

CHAPTER 1. INTRODUCTION

1.1 Nucleic acids relevance in early life and modern life

Nucleic acids (NA) are the molecular wheelhouse of living systems and are often described as the most important biomolecule present in living organisms due to their ability to store and transmit genetic information (1). As a result of their ubiquitous role in cells today, it is no surprise that they are often cited as one of the first biopolymers (2) that were present on the prebiotic earth ~3.5 bya.

In the origins of life field, an essential feature of life is the ability to not only store genetic information, but to also replicate this information and carry out stored functions in order to leave descendants (3,4). In living systems today nucleic acid replication is often a complex process involving enzymes such as helicases, primases, polymerases, topoisomerase and ligases. Helicases separate the double strands into the single strands to allow copying of each strand. Primases make the primer which is required for DNA synthesis to begin. DNA polymerases then extend this primer by the addition of one nucleotide at a time. Finally, the DNA ligase seals the nick in the phosphodiester bond formed during replication. RNA is then produced from DNA strands by a process known as transcription using RNA polymerases. Aside from the use of enzymes in nucleic acid replication today, tri-phosphates (e.g. ATPs), typically generated *in-situ*, are often used as the activated intermediate by which a phosphodiester bond is formed.

The ability of RNA to transfer information and catalyze reactions, aided by the discovery of ribozymes, has led some to postulate that RNA was the first biopolymer on

the early earth, commonly referred to as the RNA world hypothesis (5-8). In the absence of the enzymes necessary for synthesis on the prebiotic earth, the question of how RNA could have replicated remains a mystery. Although extensive work has been done to understand and prove the RNA world hypothesis (5-8), the many years of research has yet to produce a prebiotic plausible path to RNA or an efficient mechanism for replication without the use of enzymes. Some of the challenges of the RNA world hypothesis includes the strand inhibition problem, which occurs due to the high melting temperature of formed duplex products, and the problem of regioselectivity – by which newly formed linkages possess a 2´- 5´ bond rather than the canonical 3´- 5´ bond.

There are two major pathways through which RNA replication could have occurred on the prebiotic earth, either through self-condensation of monomers or templated directed synthesis (9). In template-directed synthesis, short RNA strands (oligomers) bind to single stranded regions of the templates and become "stitched" together to make a copy of the template (10). The "stitching" of oligomers is described as ligation. In the absence of ligase enzymes on the prebiotic earth, the low reaction yields encountered in ***non-enzymatic ligation reactions*** poses a major limitation to non-enzymatic replication. Other limitations such as the strand inhibition problem have been addressed by others, including work in the Hud group which focused on prebiotic pathways to multiple rounds of replication (11,12). The development of a true non-enzymatic replication cycle is dependent on the development of sustainable non-enzymatic ligation reaction systems. This thesis will focus on addressing challenges with non-enzymatic ligation.

1.2 Non-enzymatic ligation

1.2.1 Influence of non-enzymatic ligation on prebiotic chemistry and nanotechnology

Non-enzymatic ligation in prebiotic chemistry currently occurs through the use of activated species to form the phosphodiester bond. Activated species act as a better leaving group than di-phosphates, which occur during enzymatic replication with biologically produced tri-phosphate intermediates. The primers and substrates to be ligated are typically activated using organocatalysts such as imidazole species as in the case of Szostak and coworkers (13-15), 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) as in the case of Orgel and von Kiedrowski (16,17), or by a combination of both methods as used by Richert and co-workers (18-21). Organocatalysts such as imidazole often require intricate synthetic routes that pose a challenge to prebiotic chemists seeking to develop a robust and simple pathway for oligonucleotide synthesis.

More recently, there has been a large focus on ligation through an intermediate cyclic phosphate, similar to reactions in naturally occurring ribozymes (22-25). Additionally, ligation through a cyclic phosphate pathway is attractive because cyclic phosphates can occur in RNA through transesterification and cleavage of single stranded RNAs (26,27) and have also been used in nucleotide synthesis pathways (22,28-31). In synthetic RNA studies, cyclic phosphates can be generated by incubation with EDC (25). Although EDC is not prebiotically plausible, it serves as a shortcut in cyclic phosphate ligation, given that the cleavage step which reveals a cyclic phosphate in naturally occurring RNA is often slow (26,27). Building on past work, over the last few years, a plausible prebiotic route to the formation of cyclic phosphates has emerged (31,32) using

diamidophosphate (DAP), which could potentially have a large impact in non-enzymatic ligation studies, as demonstrated by Mutschler and coworkers (33).

In addition to prebiotic chemistry, non-enzymatic ligation can also aid in the development of DNA nanotechnology, since this process can allow for the ligation of DNA duplexes which contain non-Watson-Crick base pairs, non-duplex structures, and enzyme-inaccessible regions (34-36). For such applications non-enzymatic chemical ligation offers potential advantages over enzymatic ligation, as well as lower cost, which could be important for large-scale reactions. Current methods for DNA ligation occur through the use of click chemistry (37-40), the attachment of phosphorothioate groups to the oligonucleotides (41), and by the activation of the phosphate group using chemical activating agents such as cyanogen bromide (42,43), cyanoimidazole (44), and water-soluble carbodiimides (EDC) (42,45,46).

1.2.2 Challenges associated with current non-enzymatic ligation methods

In order to develop an assay for non-enzymatic ligation of RNA, characterizing ligation in a well-studied model system like DNA is beneficial. However, despite the numerous examples of non-enzymatic DNA ligation methods (47), there are still several challenges that have prevented widespread use of the DNA ligation methods. Some of these obstacles include the rapid hydrolysis of cyanogen bromide, a commonly used activating agent, and the incomplete ligations reported at even high concentrations of this highly toxic reagent. In addition, the wide variety of ligation yields ranging from as low as 30% to as high as 95% when cyanogen bromide was used as an activating agent with seemingly comparable DNA structures (42) makes it difficult to further develop this reagent as an

activating agent. Cyanoimidazole, a closely related activating agent to cyanogen bromide, allows for longer reaction times than cyanogen bromide, but has also shown limitations for achieving quantitative ligation (44).

A more benign type of activation agent is water-soluble carbodiimides (EDC), which are now a common choice of activating agents for chemical ligation in both DNA and RNA systems. In DNA, EDC activates the phosphate prior to ligation, whereas in 3′-phosphate RNA, EDC leads to the formation of the cyclic phosphate intermediate used in ligation. Previous reports have shown that DNA ligation using water-soluble carbodiimides can approach 95%, after 6 days (46,48). However, most reported yields also fall far short of quantitative ligation and are severely impacted by side reactions (49). EDC is also widely used in various synthesis reactions because of its ease of accessibility and high solubility in water. Because most reported non-enzymatic ligation reactions were conducted at low temperatures (e.g., 0 °C) it is not clear whether reaction yields or rates would be improved by increasing the reaction temperature.

It is possible that the same challenges with DNA non-enzymatic ligation will also apply to RNA ligation. Nonetheless, in the case of RNA, an additional challenge that will occur for the 3´-phosphate system is the low reactivity often encountered with the cyclic phosphate intermediate. These low reaction yields are typically ~10% (50,51), which is nine times less than the average yields for DNA non-enzymatic ligation. Given that most RNA cyclic phosphate studies have involved both the cleavage and ligation step, and given the slow rate of cleavage of the RNA strand to reveal the cyclic phosphate (26,27), the maximum attainable yield of RNA cyclic phosphate ligation is not clear.

Therefore, this thesis will focus on the optimization of non-enzymatic DNA and RNA ligation using EDC. To our knowledge there are no published studies that simultaneously explored, in a systematic manner, the combined effects of reaction temperature, template-substrate complex stability, and deleterious substrate modification on ligation product yields. The investigation of a ligation system with these variables could provide useful information for designing nucleic acid chemical ligation reactions that achieve higher yields and more convenient reaction protocols, such as shorter reaction times and fewer buffer additives. In addition, such a system could also be valuable for determining the possibility of the cyclic phosphate to generate longer strands of RNA in the prebiotic world.

The first chapters of this Thesis (Chapters 2 and 3) use a simple bimolecular oligonucleotide framework to investigate non-enzymatic DNA ligation. In Chapter 2, the effect of template and substrate stability is studied in order to obtain a correlation between equilibrium assembly and ligation yields. In Chapter 3, two primary substrates are used to investigate the relationship between increasing the reaction temperature in order to achieve high yields, while keeping the temperatures low enough to prevent EDC hydrolysis and side product formation. Using the optimized template substrate systems from Chapters 2 and 3, Chapter 4 investigates the effect of intrinsic nucleic acid features such as phosphate positions, ligation site sequence context, and sugar identity on ligation kinetics and yields. Finally, in Chapter 5, having optimized the design of the ligating system for stability and reaction temperature, these results will be extended to an RNA system. In particular, the sequence space at the ligation nick of a tri-molecular system is investigated to determine

conditions under which high ligation yields can be obtained through a cyclic phosphate intermediate.

CHAPTER 2. EFFECT OF THERMODYNAMIC EQUILIBRIUM ASSEMBLY ON DNA LIGATION KINETICS[1]

2.1 Introduction

One of the challenges of utilizing chemical ligation, particularly in the ligation of DNA and RNA strands, is the considerably low yields and long reaction times obtained compared to enzymatic reactions (42,46,48). It is not uncommon for typical DNA ligation yields to approach 95% after six-days at a reaction temperature when using water-soluble carbodiimides such as 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). Whereas, enzymatic ligation reactions, with either the commonly used T4 DNA ligase (for DNA strands) or T4 RNA Ligase 2 (for RNA strands) often result in > 99 % yields in ninety mins at the specified reaction temperature (43,52-56).

In order for chemical ligation techniques to compete with enzymatic ligation, there needs to be a systematic study to identify the different factors responsible for achieving high ligation yields. Given the complex nature of the reactivity of EDC and the different steps in chemical ligation, it is paramount to understand how the interplay of these factors will result in high yielding systems.

[1] Adapted from Obianyor, C., Newnam, G., Clifton, B. E., Grover, M. A., & Hud, N. V. (2020). "Towards Efficient Nonenzymatic DNA Ligation: Comparing Key Parameters for Maximizing Ligation Rates and Yields with Carbodiimide Activation." *ChemBioChem*, *21*(23), 3359-3370, with permission from WILEY.

Unlike enzymatic reactions, reactions in prebiotic chemistry often require the use of a template to assemble either mononucleotides for polymerization (57-60) or short substrates for subsequent ligation. In polymerization reactions, to drive assembly and increase the propensity for a nucleotide (nt) to find its pairing partners high concentrations of nucleotides are used, often in mM amounts. For example, in the ligation of dimers and trimers, Richert and coworkers (19,21) observed a ten-fold difference between the optimum dimer and trimer concentrations for ligation. Therefore, in order to understand how the difference in stability of a particular substrate-template complex might affect chemical ligation, we designed equilibrium and ligation studies using a simple bimolecular system. Given that a template is not often required in enzymatic ligation reactions, addressing this challenge is unique to non-enzymatic ligation and will lend insight into the low yields' characteristic of these types of reactions.

Results in this chapter will demonstrate the complex relationship between achieving a highly stable substrate-template complex and obtaining high chemical ligation yields.

2.2 Simple model of ligation reaction

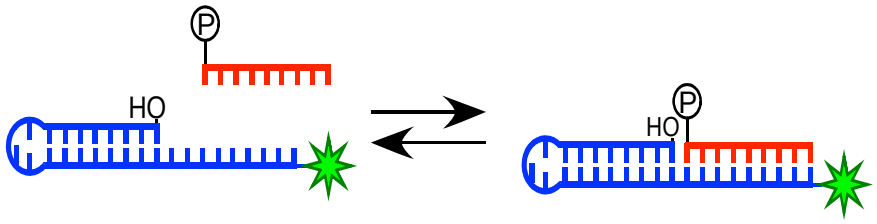

Figure 1. Simple illustration of the ligation system to demonstrate binding of substrate (red strand) to the hairpin template (blue strand).

The ligation test system is based on 28-nt hairpin-containing oligonucleotides that are labeled on their 3´ ends with 6-FAM (Fluorescein). These oligonucleotides, referred to throughout simply as "hairpins," also contain a 9-nt single-stranded overhang that serves as a substrate-binding site for template-directed ligation (Figure 1). Given that a 5-nt substrate was the shortest length that could be synthesized commercially from the supplier (IDT), we investigated the stability of different lengths of the oligonucleotides ranging from a 5-nt substrate (which would be partially complementary to the template) to a 9-nt substrate (fully complementary to the template). For the investigations reported in this chapter the phosphate group at the nick site of the ligation assembly was a 3´-phosphate.

2.3 Melting temperatures of substrate and hairpin

In order to determine the number of substrates bound to the template, the first step is to determine the melting temperature of the substrate from the template i.e. at what temperature is half of the substrate un-bound from the template. Binding of the substrate to the template can occur in two forms, either through Watson-Crick base pairing or through the ability of the substrate to remain in the double helix as a result of stacking interactions, in the case of mismatches. In the systems discussed in this chapter, the substrates are base-paired, therefore, the unfolding of the substrate from the template is primarily through the unraveling of the single stranded regions from the double helix.

UV-vis spectroscopy was used to conduct thermal denaturation studies, by monitoring the change in absorbance at 260 nm over temperature. In our system depicted in Figure 1, it was expected that there would be two melting temperatures (T_m), one for the substrate unfolding from the template, and the other from the unfolding of the hairpin. The

results of the thermal denaturation studies are shown in Figure 2 and summarized in Table 1.

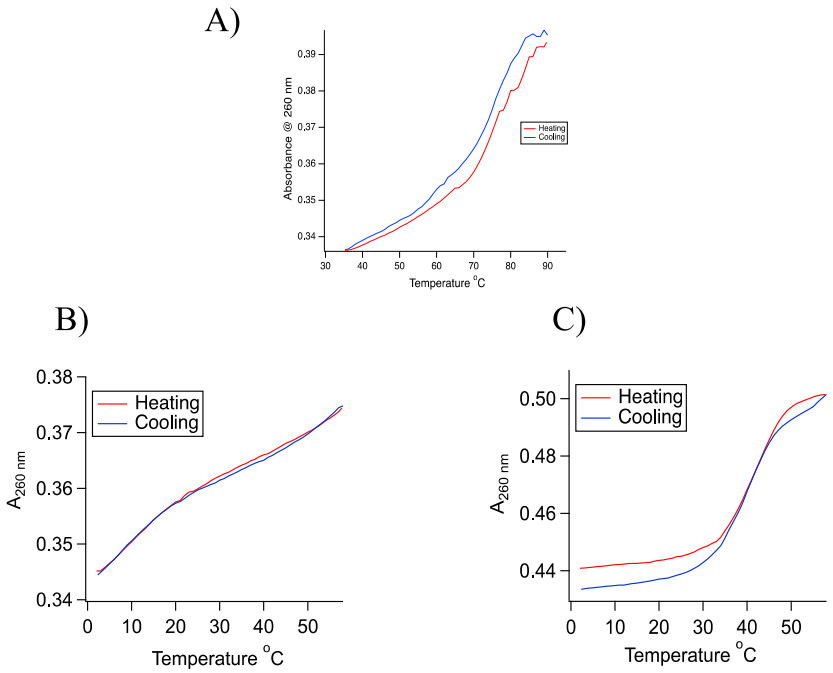

Figure 2. Determination of melting temperature (*T_m*) of the substrate-hairpin assembly by monitoring absorption at 260 nm as a function of temperature. (A) Melting of the hairpin template at a T_m of 82 °C (B) The 9-mer dissociates from the hairpin at a T_m of 45 °C. (C) No clear transition of the 5-mer hairpin dissociation can be observed over the temperature range of 2 °C and above.

Table 1. Summary of the melting temperatures of the hairpin template and substrate used in this study. N/A indicates data which could not be determined from observed spectra.

Oligonucleotide type	**Melting temperature** (^{o}C)
Hairpin	82 ^{o}C
Hairpin + 5-mer substrate	N/A
Hairpin + 6-mer substrate	N/A
Hairpin + 7-mer substrate	28
Hairpin + 8-mer substrate	38
Hairpin + 9-mer substrate	45

The melting temperature of the hairpin template was found to be 82 ^{o}C. As the length of the substrate was increased from 5-nt to 9-nt an increase in T_m is observed, as expected. It should be noted that only one transition is shown in Figures 2B and 2C in order to highlight the substrate unfolding, however, the unfolding of the hairpin is still occurring in all cases. The T_m of the 5-mer and 6-mer could not be experimentally determined despite changing the ionic strength of the buffer and the concentration of the substrates but are predicted to be around 6 ^{o}C and 20 ^{o}C respectively. The large difference between the T_m of the 9-mer and 5-mer substrates suggests that the reactivity of the ligation at temperatures close to room temperature and below will vary vastly. This difference in melting temperatures also suggests that the number of substrates bound at these temperatures will be dependent on reaction temperature and is discussed further in the next section.

2.4 Mathematical prediction of amount of bound substrate

As discussed earlier, non-enzymatic ligation chemistry is most likely dependent on the number of bound substrates available for the formation of the phosphodiester bond. In order to determine the equilibrium distribution from the number of bound substrates, the Van't Hoff Equation (Equation 1) is used to relate the melting temperatures with equilibrium constant.

In particular, we sought a set of substrate oligonucleotides that would be fully bound to the hairpin at and below 25 °C, and at least partially associated at 37 °C, the highest temperature used for our ligation reactions. A set of shorter substrates was also desired that would be partially associated with the hairpin at the lowest reaction temperature, 4 °C, but primarily or even completely dissociated at 25 °C and 37 °C. Based on experimental and predicted melting temperature (T_m) values, 9-mer and 5-mer substrates were selected and the number of bound substrates calculated.

$$\ln \frac{K_{eq,T_m}}{K_{eq,T}} = \frac{\Delta H^o}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right) \qquad (1)$$

where $\frac{K_{eq,T_m}}{K_{eq,T}}$ = the ratio between the equilibrium constant at the melting temperature (T_m) and the constant found at the different reaction temperatures (T). ΔH^o = specific enthalpy of the reaction, and R = molar gas constant.

Since the T_m of the 5-mer could not be measured, the "MELTING" program (61) based on work by Santa Lucia et. al (62) was employed. The predicted T_m of the 5-mer was -4 ^oC and 6 ^oC at an excess concentration of 1.5x and 10x respectively, while the predicted T_m of the 9-mer was 35 ^oC and 45 ^oC at an excess concentration of 1.5x and 10x respectively. For comparison, the experimental T_m for the 9-mer was found to be 45 ^oC at 1.5x substrate excess, while the predicted T_m under the same conditions was found to be 45 ^oC. The conditions between the experimental melts and the predicted melts were slightly different since the buffer for the predicted melts was assumed to be 5 mM $MgCl_2$, Tris pH 7.0, which is different from the ligation buffer conditions of 5 mM $MnCl_2$, MES pH 6.0 in which the experimental melts were calculated. Therefore, the discrepancy between the predicted and experimental T_m of the 9-mer is expected.

Using the predicted enthalpy from the "MELTING" program (i.e. - 261,668 J/mol for the 9-mer, and -136,686 J/mol for the 5-mer), the value of $\frac{K_{eq,T_m}}{K_{eq,T}}$ was found at the different reaction temperatures. Afterwards, this ratio was used to find the fraction of substrates bound at different reaction temperatures for different excess amounts, according to Equation 2 and Equation 3.

$$K_{conc,T} = \frac{x}{(A-x)(B-x)} \tag{2}$$

where x = concentration of hybridized substrate, = 0.5 at T_m

A = Initial hairpin concentration

B = Initial substrate concentration

$$\frac{K_{eq,Tm}}{K_{eq,T}}(pred) = \frac{K_{conc,\,Tm}}{\dfrac{x_T}{(A - x_T)(B - x_T)}} \qquad (3)$$

$K_{conc,T}$ is found at the T_m and used to predict the fraction of hybridized substrate x_T at different temperatures, according to Equation 3. These results are summarized in Table 2.

Table 2. Predicted fraction of hybridized substrates at different reaction temperatures for a solution containing a substrate:hairpin ratio of 10:1 and 1.5:1. These values are predicted for a solution containing 5mM $MgCl_2$ and 100mM Tris buffer.

	5-mer		9-mer	
Excess substrate	1.5 x	10x	1.5 x	10x
Reaction Temperature (°C)	Hybridized substrate (x)	Hybridized substrate (x)	Hybridized substrate (x)	Hybridized substrate (x)
4	0.18	0.6	1	1
25	4.0×10^{-3}	0.023	0.94	0.99
37	4.6×10^{-4}	2.8×10^{-3}	0.36	0.86

Our findings indicate that 100% of the hairpin in the 9-mer ligation reactions with a 1.5x excess substrate had bound substrate at 4 °C and 25 °C substrate, and approximately 90% at 37 °C. In contrast, under those same conditions, around 60% of the hairpin is predicted to have the 5-mer substrate bound at 4 °C, approximately 2% at 25 °C, and less than 0.1% at 37 ºC. These temperature-dependent substrate-hairpin complex stabilities illustrate the direct impact that changes in ligation complex stability will have on ligation reaction rates over a relatively small temperature range (e.g. 25 °C to 37 °C).

It is important to note that these estimates for equilibrium constants are based on un-activated 9-mer and 5-mer substrates. EDC has been shown to stabilize pre-ligation DNA duplexes (49), so it is possible that the T_m values for the binding of the activated 9-mer and 5-mer substrates with the hairpin will be different for EDC-activated substrates.

2.5 Kinetics of ligation for the 5-mer and 9-mer

Since ligation reactions are not spontaneous, a chemical activating agent needs to be used for ligation to occur. For all reactions shown herein, the water-soluble condensing agent EDC was used for in situ phosphate activation. Preliminary studies revealed that buffer conditions previously optimized for DNA ligation with cyanoimidazole activation (63) were also near optimal for ligation with EDC activation, and provided near quantitative ligation for some ligation systems. In general, EDC activates the 3´-phosphate of the substrate, thus allowing the phosphate to become a good leaving group upon attack by the nearby 5´-OH nucleophile of the hairpin template, as depicted in Figure 3.

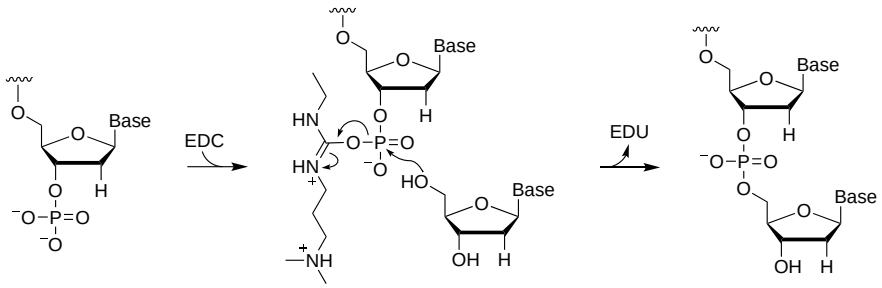

Figure 3. Key steps in the reaction mechanism for 3´-phosphate substrate activation by EDC and subsequent ligation.

In any reaction set, EDC was added to pre-mixed solutions containing hairpin and substrate oligonucleotides, and the reaction tubes were immediately moved to one of three incubation temperatures (4 °C, 25 °C, or 37 °C). Given that the half-life of EDC in solution is around 16 hr at room temperature (49,64), we were particularly interested in identifying reaction conditions that would result in full ligation within 24 hr, or 48 hr if complete ligation within 24 hr was not possible.

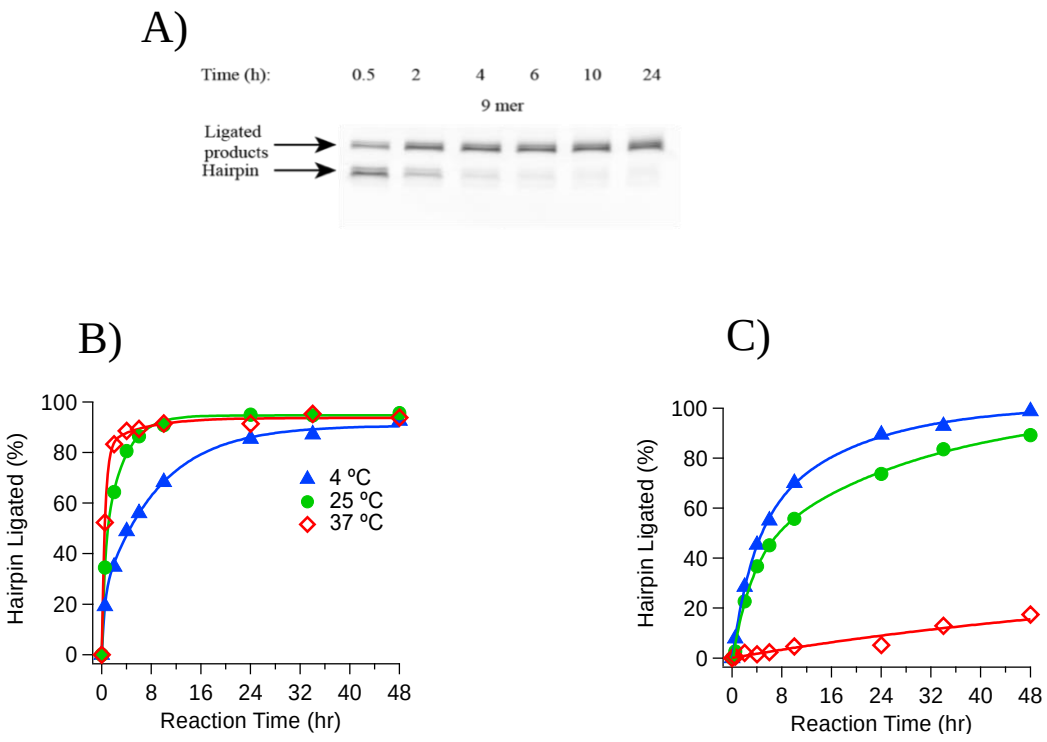

Figure 4. Kinetic measurements of chemical ligation at different temperatures. A) Sample gel electrophoresis image of kinetic studies conducted at 25 °C for the 3´-phosphate 9-mer substrate. Buffer conditions: 5 mM $MnCl_2$, 100 mM MES, pH 6.0, and 250 mM EDC. B) Kinetic data for hairpin ligation with the 3´-phosphate 9-mer substrate. C) Kinetic data for hairpin ligation with the 3´-phosphate 5-mer substrate. 1.3 µM hairpin and 13 µM of substrate DNA oligonucleotides were used for all reactions shown. Data points were obtained by scanning of polyacrylamide gels, such as the example shown in Panel A (See Materials and Methods for additional information). The lines are double exponential fits to help guide the eye through the data points as well as to provide a visual comparison of relative initial reaction rates and the time at which each set of reactions reached 50% yield. EDC hydrolysis and side reactions that reduce the concentration of ligatable substrates and hairpins are likely reasons that single exponential fits were not possible.

In Figure 4 ligation results are presented for the ligation of a particular 28-nt hairpin with its complementary 9-mer and 5-mer substrates as a function of time and temperature. For this set of reactions, the substrate oligonucleotides were present in a 10:1 relative molar concentration to the hairpin (i.e., 13 µM substrate, 1.3 µM hairpin) to ensure that the

substrate oligonucleotide would not be a limiting reagent of the reaction, a point addressed in detail in Section 2.6.

The ligation kinetics and yields observed for reactions carried out at 4 °C are quite similar for the 9-mer and 5-mer substrates. Both systems reached 50% ligation yield after reacting for 5 hr, around 80% yield at 24 hr, and over 90% yield at 48 hr (Figure 4B and C). Increasing the temperature of the 9-mer substrate reactions to 25 °C and 37 °C resulted in a substantial increase in ligation rates, with 50% ligation being reached within 1 hr, and maximum ligation yields of around 90% for both temperatures being achieved near the 10 hr time point. In contrast, increasing the temperature of the 5-mer substrate reactions to 25 °C caused a minor decrease in the ligation rate and in the ligation yield measured at 48 hr. At 37 °C the ligation rate of the 5-mer substrate reaction was decreased so drastically that only 17% yield was measured at the 48 hr time point (Figure 4C).

The opposite response of the 9-mer and 5-mer substrate ligation reactions to increased temperature illustrates the interplay of two fundamental factors (stability of substrate/template complex and chemical ligation step) that govern ligation rates in our ligation test system and, presumably, in many published studies involving DNA chemical ligation. For the 9-mer substrate reactions, the higher ligation rates at 25 °C and 37 °C can be attributed to an increase in the rate of the chemical step of ligation with temperature, given the high amount of bound hairpin and substrates as discussed in Section 2.4. The rates of the chemical steps of the 5-mer substrate reactions will also increase with temperature. Thus, the observed *decrease* in 5-mer substrate ligation rates at 25 °C and 37

°C is likely due to a decreased fraction of hairpins with bound 5-mer substrates at higher temperatures.

2.6 Impact of substrate/hairpin ratio on ligated products

The reaction kinetics studies presented above all had a substrate:hairpin molar ratio of 10:1. This ratio was chosen based on preliminary studies which showed that substrate:hairpin molar ratios closer to 1:1 resulted in incomplete ligation at 24 hr for the 9-mer and the 5-mer substrates at all three reaction temperatures. Given the difference in stability between our 9-mer and 5-mer substrate-hairpin complexes, it was not clear if, or to what degree, incomplete ligation was limited by reaction kinetics or by deleterious oligonucleotide modification during the course of the reactions, to be discussed further in Chapter 3 (21,63).

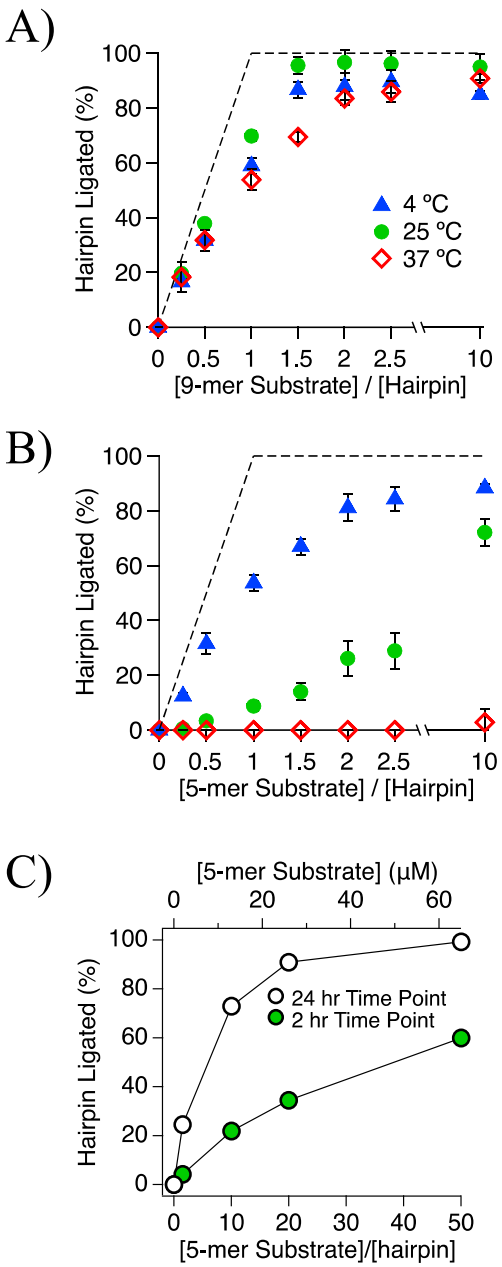

Figure 5. Effects of substrate-hairpin ratios on ligation efficiency for 3´-phosphate reactions. (A) Product formation for the 9-mer substrate as a function of relative molar concentrations of substrate to hairpin, after 24 hr reaction time. (B) Product formation for the 5-mer substrate as a function of relative molar concentrations of substrate to hairpin, after 24 hr reaction time. (C) Product formation for the 5-mer substrate comparing the 2 hr and 24 hr yields at 25 °C as a function of relative substrate:hairpin molar concentrations. Dashed lines in panels A and B indicate maximum possible yields for reactions as a function of substrate:hairpin ratios. All reactions contained 1.3 µM hairpin and were carried out in a buffer containing 5 mM $MnCl_2$, 100 mM MES pH 6.0, and 250 mM EDC. Error bars represent the range of

values measured for three separate experiments. Yields were obtained as described in the caption to Figure 4.

Plots of ligation yields at the 24 hr reaction time point for various substrate:hairpin ratios are presented in Figure 5 for the 9-mer and 5-mer ligation systems. These plots show that reactions with the 9-mer system at 4 °C and 25 °C achieve maximal 24-hr yield (as indicated by plateau) for substrate:hairpin ratios of 1.5:1 and greater (Figure 5). Higher ratios are required for the 37 °C reaction to reach the same level. For reactions with the 5-mer substrate, the 25 °C and 37 °C reactions do not reach their respective maximum possible 24-hr yields even at the 10:1 substrate:hairpin ratio. The 5-mer substrate reactions at 4 °C appear to be approaching a comparable 24-hr yield plateau at the 10:1 ratio, but a significantly higher concentration is needed to reach maximum yield.

In the plots shown in Figure 5A and 5B the diagonal and horizontal dashed lines indicate where data points would be located if the hairpins molecules were quantitatively ligated to the substrate molecules within the 24 hr reaction time. The diagonal line corresponds to the region of the plot where the molar concentration of the substrate oligonucleotides is less than that of the hairpin oligonucleotides. To ensure that the consistently less than possible ligation yields at substrate:hairpin ratios near and above 1:1 were not due an underestimation of hairpin concentration, we determined the extinction coefficient by experimental means and verified the manufacturer provided extinction to be correct (see section 2.9.6 for more details) (65). Thus, the data points of all experiments fall below the dashed lines because for all reactions studied hairpin ligation with substrates present is not complete within 24 hr.

Reactions with various substrate to hairpin ratios carried out at different temperatures illustrate that three factors primarily limit our ligation test system from reaching quantitative yields. The 4 °C data sets show how low reaction temperature can render the chemical step of the reaction too slow for full ligation to occur even though hairpin-substrate association is strongly favored. As noted above, at 4 °C hairpins in the 9-mer reactions are fully associated with substrates, and there is also significant association for the 5-mer substrate reactions at 4 °C. Nevertheless, no reactions carried out at 4 °C reached 100% yield of ligated hairpin by 24 hr, regardless of there being even 10-fold excess substrate.

The second reason that some reactions presented in Figure 5A and 5B do not reach their maximum possible yields is that substrate *and* hairpin oligonucleotides are being rendered unligatable by side reactions with rates that are comparable to the ligation reaction rate. The impact of this limitation is illustrated by the 9-mer substrate reactions run at 37 °C with a substrate:hairpin ratio of 10:1, which were shown above to reach their maximum yield in about 12 hr (Figure 4B). The fact that these reactions fall short of complete hairpin ligation, despite exhibiting rapid initial kinetics and containing 10-fold more substrate than hairpin, suggested to us a significant loss of hairpin and substrates to side reactions at 37 °C. The effect of side reactions will be further explored in Chapter 3.

The third factor contributing to limited ligation within 24 hr is minimal equilibrium assembly of the hairpin-substrate complex. This factor, which has been widely appreciated in previous studies of mononucleotide and short oligonucleotide template-directed polymerization, appears to be the main reason for the low yields of the 5-mer substrate

reactions run at 25 °C and 37 °C (Figure 5B). In particular, the near linear increase in the 24-hr yield of the 5-mer reactions run at 25 °C as a function of substrate concentration were consistent with observed yields being directly proportional to the concentration of assembled hairpin-substrate complex. Therefore, increasing the substrate concentration to much higher values might allow for 100% hairpin ligation. As a test of this possibility, in Figure 5C the 2 hr and 24 hr yields for this reaction are shown for much higher substrate concentrations, with substrate to hairpin ratios ranging from 1.25:1 to 50:1 for the hairpin concentration still fixed at 1.3 μM. Important observations of this set of experiments includes the 2 hr yield of the 5-mer reaction with 50x substrate being 60%, which is very close to the 2 hr yield observed for the 9-mer reaction also run at 25 °C. Thus, the 5-mer reactions at 25 °C with 50x substrate exhibit a ligation rate that is consistent with the hairpins in this reaction being fully associated with substrate. Consistent with this conclusion, the 24-hr yield for the 5-mer reaction with 50x substrate is 99% (Figure 5C), which is slightly higher than that observed for the 9-mer ligation reaction run at 25 °C with a 10:1 substrate to hairpin ratio (Figure 5A); a reaction that we are confident to have all hairpins bound to substrates.

2.7 Impact of additives on ligation kinetics

2.7.1 Organocatalysts (1-Ethyl Imidazole)

Recent study by Richert and co-workers demonstrated that 1-Ethylimidazole (1-Ethyl) greatly increased the ligation yields of RNA dinucleotides and trinucleotides in reactions with EDC as the activating agent (21). The same group had previously demonstrated that mononucleotides would undergo template-directed polymerization

when activated by EDC only if 1-Ethyl or a similar organocatalyst was present in the reaction buffer (19). In contrast, this reagent appeared to inhibit the ligation of 5´-phosphate DNA oligonucleotides within the context of a DNA origami structure (46). Given these mixed results, we decided to investigate the impact of 1-Ethyl on our DNA ligation reactions in which the phosphate at the nick site is on the substrate 3´ terminus.

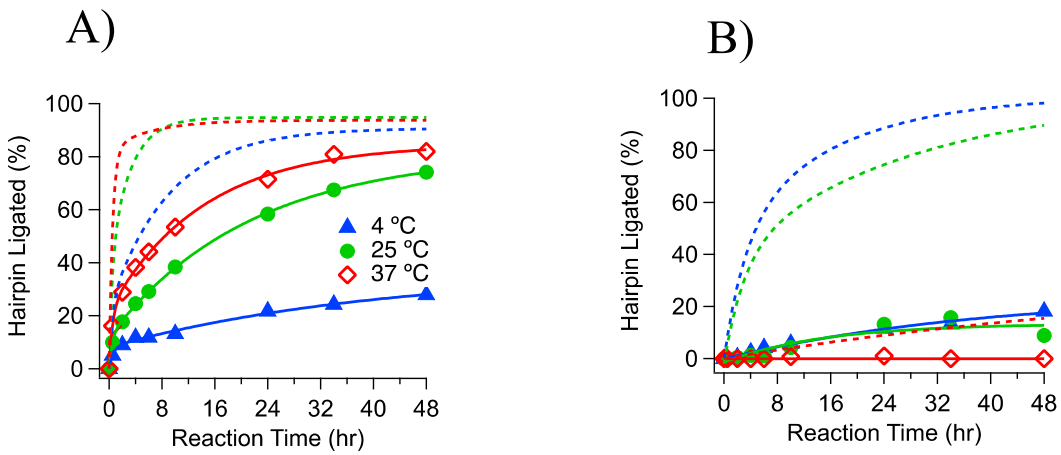

Figure 6. Effect of adding 1-Ethylimidazole (1-Ethyl) to ligation reactions. A) Kinetic data for hairpin ligation with the 3´-phosphate 9-mer substrate. B) Kinetic data for hairpin ligation with the 3´-phosphate 5-mer substrate. Markers indicate experimental results for reactions with the addition of 1-Ethyl. Solid lines are double exponential fits of data, as described in Figure 4. Dashed lines are from the data of corresponding reactions shown in Figure 4. All data points were obtained as described in Figure 1 caption. Reaction buffer: 5 mM $MnCl_2$, 100 mM MES, pH 6.0, 150 mM 1-Ethylimidazole, pH 6.0, and 250 mM EDC.

In Figure 6 we present the results from reactions that are identical to those shown in Figure 4, except in which 1-Ethyl (150 mM) was added to the ligation buffer along with EDC. As is evidenced from these plots, 1-Ethyl caused a reduction in the reaction rates for both the 9-mer and the 5-mer ligation reactions. The 9-mer substrate reactions carried out at 25 °C and 37 °C might eventually reach yields comparable to reactions without 1-Ethyl, but we estimate these levels would require approximately three times as long (Figure 6A).

This suppression of reaction rates was even more dramatic for the 5-mer-ligation system, with ligation rates and 48 hr yields falling by almost an order of magnitude (Figure 6B). Given our observation that 1-Ethyl does not enhance ligation rates or yields in our DNA system, we did not include 1-Ethyl in any other reaction reported in this study.

2.7.2 Monovalent and divalent salts

Reaction buffer ionic strength and ionic species are other possible modulators of ligation rates and yields. Therefore, we also investigated the effects of monovalent and divalent salts on our ligation reactions. Given the different reports (26,66) on the effect of divalent salts on ligation, we sought specifically to test whether our system would perform better with Mg^{2+} or Mn^{2+} ions present (Figure 7). Our studies confirmed that there was no appreciable difference between using $MgCl_2$ or $MnCl_2$ as the divalent cation salt (Figure 7). Moreover, it was revealed that our ligation reactions could be performed without divalent cations, as the reactions in the first lane of Figure 7A have a similar amount of product formed compared to the other lanes, similar to findings reported by Braun and co-workers (49). Increasing the amount of $MnCl_2$ from 5 mM (for a standard ligation reaction buffer) to 300 mM (Figure 7B) also seemed to have little effect on the ligation rates and yields, indicating that the divalent cation concentration is saturated at 5 mM.

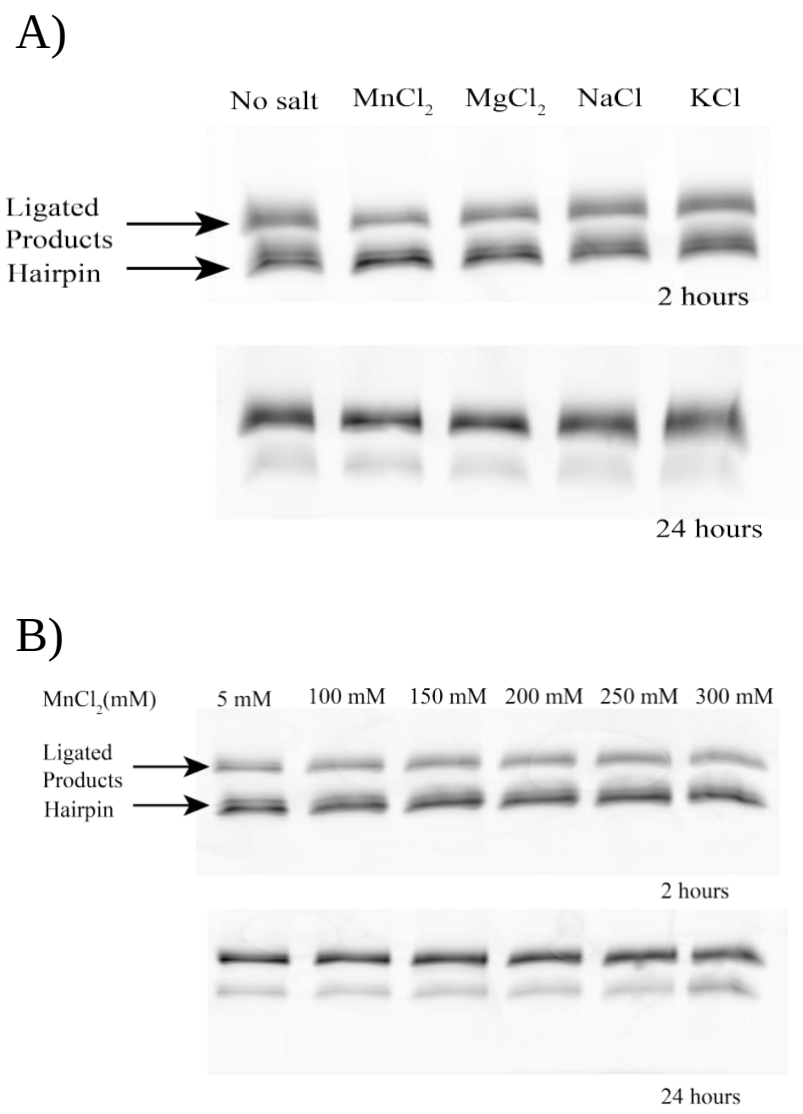

Figure 7. Gel electrophoresis images showing the effect of different salts on 3´-phosphate 9-mer ligation. (A) 5 mM concentration of different divalent and monovalent salts were added to the ligation buffer. (B). The concentration of $MnCl_2$ was varied in the ligation buffer. For all experiments, the reaction was monitored after 2 hours and 24 hours respectively at 4 °C and the hairpin concentration was 1.3 µM while the substrate concentration was 2 µM. Buffer conditions: 100 mM MES, pH 6.0, and 250 mM EDC.

There is a contrast between yields obtained from adding monovalent salts and divalent salts, as observed in Figure 7A. The initial rate of ligation (estimated by products formed after 2 hours) is higher with the addition of monovalent salts compared to divalent

salt yields. Driven by these results, kinetic studies were performed to investigate the overall effect of adding a 100 mM NaCl to the standard ligation reaction buffer (Figure 8).

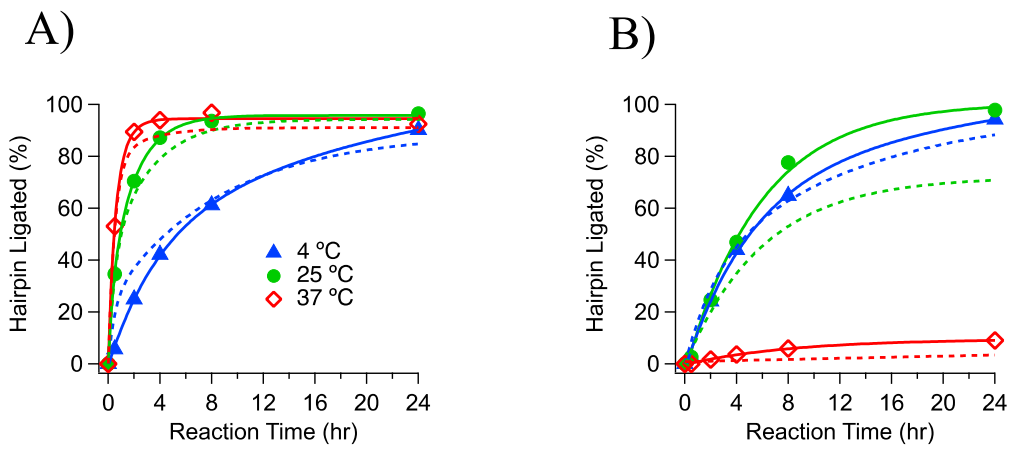

Figure 8. Impact of adding 100 mM NaCl to the ligation reaction buffer. (A) Kinetic data for hairpin ligation with the 3´-phosphate 9-mer substrate. (B) Kinetic data for hairpin ligation with the 3´-phosphate 5-mer substrate. Markers and solid lines are yields observed with 100 mM NaCl added to the ligation buffer. Dashed lines are double exponential fits of data for corresponding reactions in standard reaction buffer without the addition of NaCl. Standard reaction buffer: 5 mM $MnCl_2$, 100 mM NaCl, 100 mM MES, pH 6.0, and 250 mM EDC. Reaction temperatures as indicated in A.

The addition of 100 mM NaCl had minimal impact on the 9-mer substrate ligation kinetics, shown in Figure 8A. The initial rates and yields are almost within experimental error of the results obtained without the addition of NaCl. Similarly, the ligation reaction with the 5-mer substrate carried out at 4 °C exhibited negligible change in the reaction kinetics and overall yields (Figure 8B).

Significantly, reactions with the 5-mer substrate carried out at 25 °C exhibited an appreciable increase in ligation rates with added NaCl. Specifically, the ligation rates at 25 °C increased to approximately those measured at 4 °C with or without NaCl. On the other hand, even with the added NaCl the 37 °C reaction ligation rates remained low (i.e., <10%

at the 24 hr time point). The increase in the rate of ligation of the 5-mer at 25 °C is to be expected since higher salt concentrations typically increase the T_m of duplex oligonucleotides (62,67). In summary, given the low T_m of 5-mer compared to the 9-mer substrate, it is unsurprising that the addition of monovalent salts had a higher impact on the ligation kinetics of the former compared to the latter.

2.8 Conclusion

In this chapter, we laid out the framework for characterizing a basic ligating system. Particularly, we investigated the effect of equilibrium assembly on ligation kinetics and overall yields. Thermal denaturation studies and empirical calculations confirmed that the difference in stability between the partially complementary substrate (5-mer) and the fully complementary substrate (9-mer) led to a significant difference in optimum reaction temperature for each reaction substrate. The 5-mer needed a 50x higher number of excess substrates in order to compete with the 9-mer at high reaction temperatures.

We also investigated the effect of additives on ligation kinetics. Neither the addition of 1-Ethyl nor addition NaCl appeared to increase the reaction rates. In general, the 5-mer showed optimum reactivity at 4 °C while the 9-mer showed optimum reactivity at higher temperatures of 25 °C and 37 °C. In Chapter 3, I will demonstrate other detrimental effects of high reaction temperatures on ligation yields.

2.9 Materials and Methods

2.9.1 Oligonucleotide sequences

Hairpin	5´ CAGTCACGGAACGTGACTGGACAGGAGA 3´ 6-FAM
9-mer 3´ phosphate substrate	5´ TCT CCT GTCp 3´
8-mer 3´ phosphate substrate	5´ CT CCT GTCp 3´
7-mer 3´ phosphate substrate	5´ T CCT GTCp 3´
6-mer 3´ phosphate substrate	5´ CCT GTCp 3´
5-mer 3´ phosphate substrate	5´ CT GTCp 3´

2.9.2 Melting temperature (T_m) determination

UV absorbance was used to monitor the thermal denaturation of the hairpin and complementary short oligonucleotide complex. DNA samples were prepared with 5 mM $MnCl_2$, and 100 mM MES pH 6.0, with a template hairpin concentration of 1.3 µM, and a substrate concentration of 2 µM. UV measurements were performed on 10 mm quartz cuvettes in a temperature-controlled UV-Vis spectrophotometer (Cary Agilent UV-Vis Multicell Peltier) with nitrogen flowing through the sample chamber at low temperatures. To determine T_m values, heating and cooling traces were generated for each sample by recording spectra at 260nm from 18 to 58 ^oC at intervals of 1 ^oC. T_m values were determined as described by Mergny & Lacroix (68).

2.9.3 Oligonucleotide preparation

All oligonucleotides were purchased from Integrated DNA Technologies (IDT) where they were synthesized using standard phosphoramidite methods. Upon receipt from IDT, the oligonucleotides were resuspended in 18.2 MΩ/cm water (Barnstead NanopureTM). The 3´p hairpin oligonucleotides were further purified using denaturing polyacrylamide gel electrophoresis. The 5´p hairpin oligonucleotides were prepared using T4 Polynucleotide Kinase (New England Biolabs).

2.9.4 Ligation experiments

For a standard reaction, the fluorescently-labeled hairpin oligonucleotide was present at 1.3 μM along with the substrate oligonucleotides (at concentrations ranging from 0.5x to 50x hairpin concentration), in a buffer containing 5 mM $MnCl_2$, 100 mM MES, pH 6.0, and 250 mM (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide) EDC (Sigma-Aldrich). EDC was stored at -20 ^{o}C prior to use.

Before addition of EDC, the oligonucleotides and buffer mixture were heated to 80 ^{o}C for 2 min, then quickly cooled to room temperature. Five minutes after cooling, EDC was added to the reaction to initiate the reaction, and the reaction tube was immediately moved to the specified temperature. After the indicated reaction time each reaction was quenched by the addition of an equal volume of 2x loading buffer and dye (95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, 5 mM EDTA pH 8.0). Each sample was then stored at -80 ^{o}C until analyzed by denaturing polyacrylamide gel electrophoresis.

Polyacrylamide gels (Fisher BioReagentsTM acrylamide/bis-Acrylamide 29:1, 40% solution) were 20% denaturing gels (8 M urea) run in 1x TBE (Tris, Boric Acid, and EDTA pH 8.0) buffer, 16 cm wide X 16 cm long. Gels were pre-run at 14 W and 300-400 V for at least 30 min prior to loading. Samples were run at the same conditions for 1 h. Imaging was done using a Typhoon Trio+ laser scanner (GE Healthcare) at a resolution of 50 μm and with a photomultiplier setting between 300-500. 'FAM filter' images were acquired using the 'FAM channel', which refers to 488 nm excitation and a 526 nm emission filter. Densitometry analysis was performed using utilities within the ImageJ software package (NIH).

2.9.5 Densitometry analysis of ligated products

To obtain "Hairpin % ligated products", the integrated intensity (densitometry measurement) of each gel band corresponding to a ligation product was divided by the sum of the integrated intensity of the ligation product and unreacted hairpin oligonucleotide in each lane. These % products are used to describe all ligation results.

2.9.6 Verifying manufacturer supplied extinction coefficient for the hairpin template[2]

To rule out the possibility that the hairpin concentration was being underestimated in this study, we investigated the accuracy of the nearest neighbors extinction coefficient reported by Integrated DNA Technologies for the hairpin, which is 0.307 μM^{-1} cm^{-1} and includes attenuation by 3′ 6-FAM. Given that the template forms an intramolecular structure, the well-known DNA hypochromic effect may result in an underestimation of

[2] Experiments in Section 2.9.6 were performed by collaborator Bryce E. Clifton.

the amount of hairpin in solution, which we presumed to be the limiting reagent at these ratios. Therefore, underestimating the hairpin concentration would significantly reduce observed ligation yields.

Initially, the molar concentration of the purified hairpin used in the ligation reactions was determined using an Agilent 8453 UV-Vis spectrophotometer after dilution in nuclease-free water to ~0.3 AU at 260 nm. To verify the reported extinction coefficient, two methods were used to disrupt internal stacking and base pairing for UV absorbance measurements: exonuclease digestion to the individual 5′ dNMPs and thermal denaturation of the hairpin.

a. In the exonuclease digestion method, 62 pmol of the template hairpin were diluted into 1X CutSmart buffer with 30 U T7 Exonuclease and 20 U *E. coli* Exonuclease 1 from New England Biolabs to cut from the double-stranded 5′ end and the single-stranded 3′ end, respectively. This solution (nominally 0.83 µM hairpin template) was transferred to a 10 mm quartz cuvette and covered by a layer of mineral oil and parafilm to prevent evaporation. The cuvette was incubated in the spectrophotometer at 37 ^{o}C and absorbance at 260 nm was measured after 4 and 6 hours with a 3-second integration time and 1 nm intervals. 1 µL samples were quenched at 4 and 6 hours with 4 µL 2X loading buffer and dye for denaturing gel electrophoresis, as described in the text, to ensure complete enzymatic hydrolysis. The measured absorbances were corrected with a solution containing enzymes and buffer, but no DNA. The absorbance of the hydrolyzed sample and the sum of the molar extinction coefficients of the individual 5′ dNMPs (i.e., without nearest-neighbor effect) were used to more accurately calculate

the concentration of the intact hairpin template. This experimentally determined concentration was compared to the concentration determined from the nearest-neighbors extinction coefficient using the same dilution factor of the intact hairpin template, but in pure water and at 20 ºC.

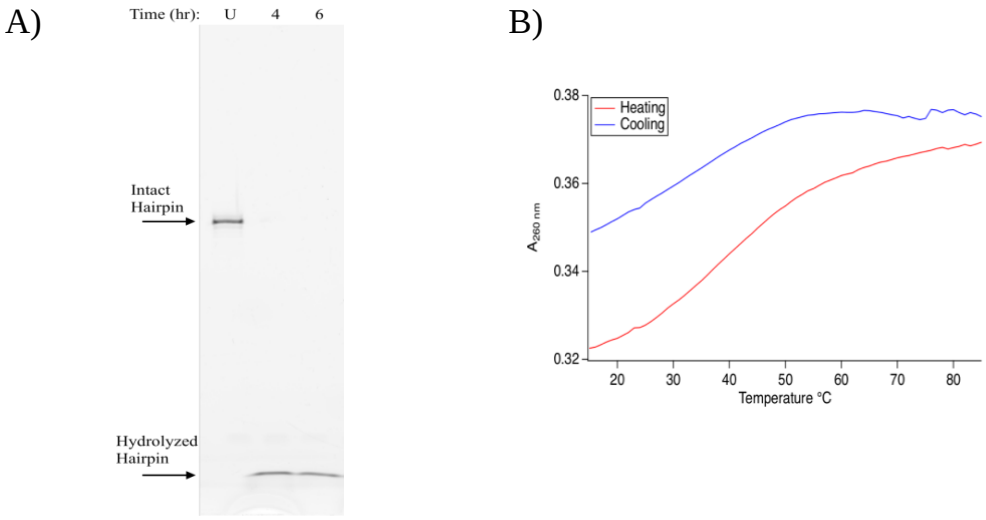

Figure 9. Verification of hairpin extinction coefficient. (A) Gel electrophoresis image of the enzymatically-hydrolyzed hairpin template. Lane "U" is the unincubated hairpin representing the intact size control. Lanes "4" and "6" are the time points (in hours) of the enzymatic digestion of the hairpin. Trace amount of the band labelled "Intact Template" is observed by 4 hours and is completely gone at 6 hours. There are no observable hydrolytic products with the exception of the single band labeled "Hydrolyzed Hairpin," which is inferred to be the 3′ terminal 6-FAM-labelled deoxynucleotide due to its fluorescence and migration being faster than the bromophenol blue dye, which is known to migrate as an 8-mer or faster. (B) Thermal denaturation of hairpin template in the absence of buffer or additional salt. Absorption is monitored at 260 nm. The weak sigmoidal character of the curve and the observed hysteresis are possibly due to the low ionic strength.

Figure 9A shows complete hydrolysis of the nuclease-treated template. The absorbances and extinction coefficients of the intact and hydrolyzed samples are reported in Table 3. The absorbance did not increase appreciably after the 4-hr measurement. The intact hairpin has a hypochromic effect that reduces the absorbance by 0.87 compared to

hydrolyzed sample. The nearest-neighbors calculation similarly reduces this extinction 0.89 times compared to the sum of extinction coefficients of the individual 5′ dNMPs with the 6-FAM extinction coefficient included. Therefore, this yields an experimentally measured micromolar extinction coefficient of 0.30 μM^{-1} cm^{-1} and a concentration of 17.00 µM, which gives a 2% difference toward a slight underestimation of hairpin in solution when using the nearest-neighbors model.

Table 3. Absorbance values and concentrations of intact and hydrolyzed hairpin template. The hydrolyzed sample is the 6-hr time point of the enzymatic digestion in Figure 9A that is apparently completely hydrolyzed.

Hairpin sample	A_{260}[a]	ε_{260}[b]	Concentration[c]
Intact in water	0.25	0.30[d]	16.66
Hydrolyzed in digestion buffer	0.29	0.34[e]	17.00

[a] Absorbance of the hairpin sample when diluted 20-fold.

[b] Extinction coefficients in units μM^{-1} cm^{-1}

[c] Concentration in units µM when corrected for 20-fold dilution

[d] Nearest-neighbor micromolar extinction coefficient provided by IDT.

[e] Sum of micromolar extinction coefficients of individual 5′ dNMPs (69).

b. In the thermal denaturation method, UV absorbance was monitored as described in Section 2.9.2 in the absence of salt or buffer, for a melting temperature range of 15 to 90 ^{o}C. This was done to replicate the conditions in which original concentration determination was done (i.e., low ionic strength conditions that do not favor secondary structure formation). The absorbance values were compared at 20 ^{o}C, at which the original quantification was measured, and 80 ^{o}C, at which the DNA should be completely denatured. In Figure 9B, a weak sigmoidal increase in absorbance is indeed seen. This weak sigmoidal character and hysteresis observed can be explained by low ionic strength. The absorbance increases by a factor of ~1.1 when comparing the room temperature and UV-Vis measurements. In other words, folding of the hairpin induces hypochromicity by a factor of ~0.91, which qualitatively agrees with the former method. Therefore, we accepted the extinction coefficient reported by IDT as sufficient for our use, suggesting incomplete ligation can be explained solely by side product formation.

CHAPTER 3. DRIVING TOWARDS HIGH YIELDING DNA LIGATING SYSTEMS[3]

3.1 Introduction

Several factors aside from equilibrium assembly of the template and substrate, discussed in Chapter 2, can also influence ligation kinetics and yields. One of these factors include the formation of unwanted side reactions, which have been reported in previous chemical ligation studies (15,18,27), but which have not been systematically characterized within these studies. Cyclization of the substrates is often reported as the major side product from EDC-based ligation reactions but other reactions such as base modifications due to reactivity of the EDC can also hinder the reactions (21,49,70). The effects of base modifications within the context of ligation reactions has not been studied and will be addressed in this Chapter, particularly given that cyclization of does not appear to limit reactivity of longer oligonucleotides similar to those used in our studies (63,71,72).

The effect of reaction temperature on ligation kinetics and yields will also be studied in our reaction system. Given that most reported non-enzymatic ligation reactions were conducted at low temperatures (e.g., 0 °C) it is not clear whether ligation yields can be improved solely by increasing the reaction temperature (42,44,45,73). Specifically, a question that will be addressed in this chapter is whether an increase in reaction

[3] Adapted from Obianyor, C., Newnam, G., Clifton, B. E., Grover, M. A., & Hud, N. V. (2020). "Towards Efficient Nonenzymatic DNA Ligation: Comparing Key Parameters for Maximizing Ligation Rates and Yields with Carbodiimide Activation." *ChemBioChem*, *21*(23), 3359-3370, with permission from WILEY.

temperature (for a fully assembled complex) will always lead to an increase in ligation yield, given the propensity to also increase side product formation.

In addition to low reaction temperatures, most reported chemical ligation studies are typically allowed to react for several days (19,21), perhaps to allow optimum time for the template and substrates to find each other. Given the reported half-life of these activating agents, and in particular the known hydrolysis of EDC (74,75), the interplay between the reactivity of the activating agents and its effect on ligation yield will also be studied.

Overall, in Chapter 3 the systematic investigation of the effects of reaction temperature, template-substrate complex stability, and substrate modification on ligation product yields will be addressed. Various experimental and mathematical modeling tools will be applied in order to decipher the key reaction parameters for an optimum ligating system. In sum, by the end of the chapter, we plan to identify reaction conditions that maximize DNA ligation yield while minimizing reaction time.

3.2 Comprehensive ligation model

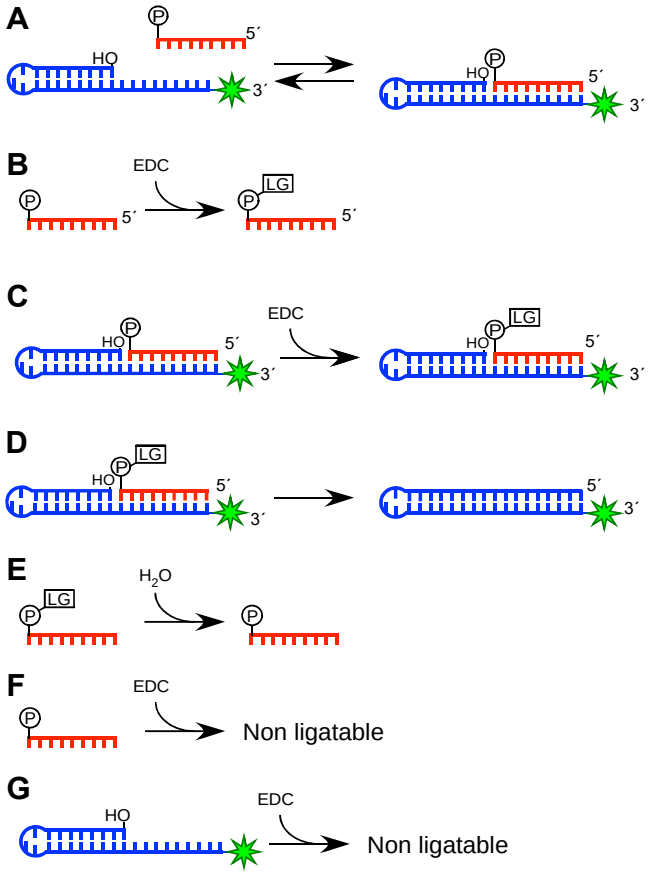

Figure 10. Illustration of the ligation test system with the equilibria and reactions that govern the overall rates and yields of ligation. A) Equilibrium of the hairpin substrate complex. B) Activation of the substrate oligonucleotide by EDC. LG is the leaving group formed upon reaction with EDC. This reaction is shown for a substrate oligonucleotide free in solution but may also occur while the substrate is bound to the hairpin. C) Activation of substrate on the hairpin substrate complex. Note that there can also be exchange of the activated substrate with solution before ligation. D) Formation of the phosphodiester bond. E) Hydrolysis of activated substrate. F) EDC-induced reaction that renders a substrate oligonucleotide non-ligatable (e.g., cyclization). G) EDC-induced reaction that renders a hairpin template unable to participate in ligation (e.g., base modification that prevents substrate binding).

Figure 10 is an illustration of the overall ligating system. In Figure 10A, the equilibrium association of the hairpin and template is shown, as discussed in Chapter 2, and is governed by an equilibrium distribution that depends on several factors, including substrate length, hairpin/substrate concentrations, ionic strength, and temperature. LG is

the leaving group formed upon reaction with EDC. Figures 10B and 10C demonstrate the activation of the phosphate with EDC. Figure 10D, shows the chemical ligation step of the substrate to the template to form the ligation product.

The last three reactions depicted in Figure 10 represent side reactions that can reduce ligations rates and yields, including hydrolysis of the EDC leaving group from an activated substrate (deactivation) (Figure 10E). EDC-induced reactions that inhibit substrates and hairpins from participating in the ligation reaction, such as cyclization in the case of the substrate, or base modification of either substrate or hairpin oligonucleotides that interfere with hairpin-substrate association (Figure 10F and 10G). The identity of the side products will be investigated in the next chapters.

3.3 Detrimental effects of side products in highly assembled complexes

The impact of side products on ligation kinetics can be more closely seen in reactions involving an equimolar/near equimolar amount of substrate and template. In Figure 11, A and B illustrate ligation kinetics for a 1.25:1 substrate:hairpin molar ratio for the 9-mer and 5-mer system respectively. While Figures 11 C and D illustrate kinetics of a 1:1 substrate:hairpin molar ratio for the 9-mer and 5-mer system respectively.

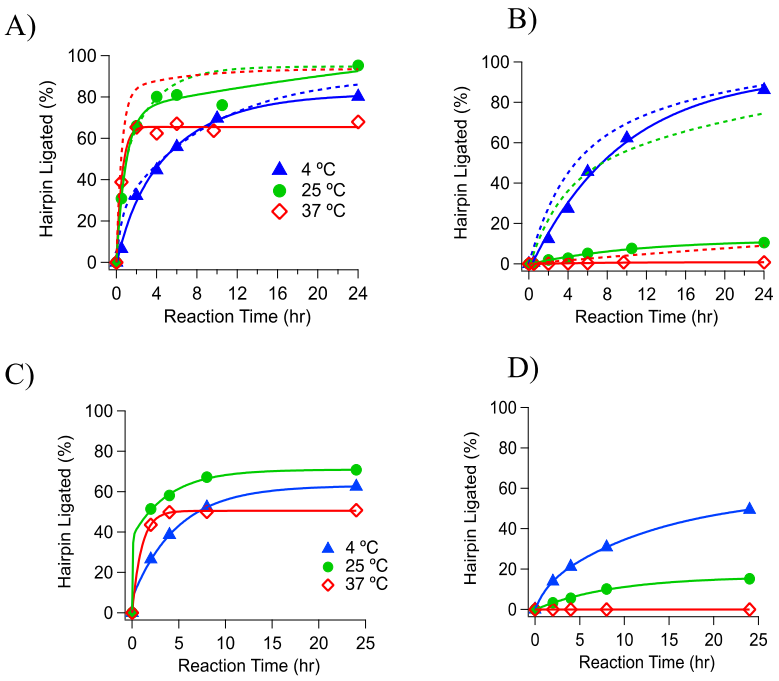

Figure 11. Kinetics of chemical ligation using near equimolar substrate and template. (A) Kinetic data for 3′-phosphate 9-mer substrate with a substrate:hairpin ratio of 1.25:1. (B) Kinetic data for the 3′-phosphate 5-mer substrate with a substrate:hairpin ratio of 1.25:1. (C) Kinetic data for 3′-phosphate 9-mer substrate with a substrate:hairpin ratio of 1:1. (D) Kinetic data for the 3′-phosphate 5-mer substrate with a substrate:hairpin ratio of 1:1. Markers are experimental data, solid lines are double exponential fit of data to guide the eye. Dashed lines are double exponential fits of data when reaction was conducted using a substrate:hairpin ratio of 10:1

The impact of side product limitation is particularly illustrated by the 9-mer substrate reactions run at 37 °C with a substrate:hairpin ratio of 10:1, which were shown above to reach their maximum yield in about 12 hr (Figure 4B). Under these same reaction conditions, Figures 11A and 11C demonstrate a high reaction rate at 37 °C but a maximum yield of only 65 %, and 50% for the case of a 1.25:1 and 1:1 molar ratio respectively regardless of reaction time. For the 5-mer reaction set, (Figures 11 B and D), there is no appreciable difference in the reaction kinetics at 37 °C for a substrate:hairpin ratio of 10:1 or 1:1, given the previous discussion on the stability of the equilibrium complex. No

conclusion can be drawn from this dataset about the effect of side product modification on the 5-mer set. Thus, there is a need for further analysis which will be described in later sections in this chapter.

The effect of side product modification at 4 °C for both the 5-mer and 9-mer is seen to a lesser degree at a substrate:hairpin ratio of 1.25:1. Given our equilibrium calculations in Chapter 2, we hypothesize that more modifications occur with an increase in reaction temperature, due to both increased side product kinetics and the fact that more substrates are unbound at higher reaction temperatures.

3.4 Identity and kinetics of substrate side product formation

To determine if oligonucleotide modification was the most limiting factor for reactions carried out at 37 °C we sought to quantify the rate and extent to which modifications are made to substrate oligonucleotides. For comparison and for completeness these studies were carried out at all three temperatures and for both the 9-mer and the 5-mer substrates. Substrate modification by EDC was monitored in the standard reaction conditions, except in the absence of the hairpin.

As expected, two primary types of modifications were identified, oligonucleotide cyclization and nucleobase modification. Based on integrated intensities of UV-monitored HPLC peaks and MS analysis, it was determined that 7% of the 5-mer substrate became cyclized after 24 hr at 4 °C. The remaining 93% showed no evidence of other modifications (Figure 12A). At higher temperatures EDC caused more and varied modifications of the 5-mer substrate. At 25 °C the amount of cyclization increased to 21% and an additional 6% of the 5-mer substrate exhibited altered HPLC retention times consistent with nucleobase

modifications, resulting in 73% of the substrate remaining unmodified after 24 hr (Figure 12A). Similar results were observed for the 5-mer substrate incubated at 37 °C for 24 hr (Figure 12A).

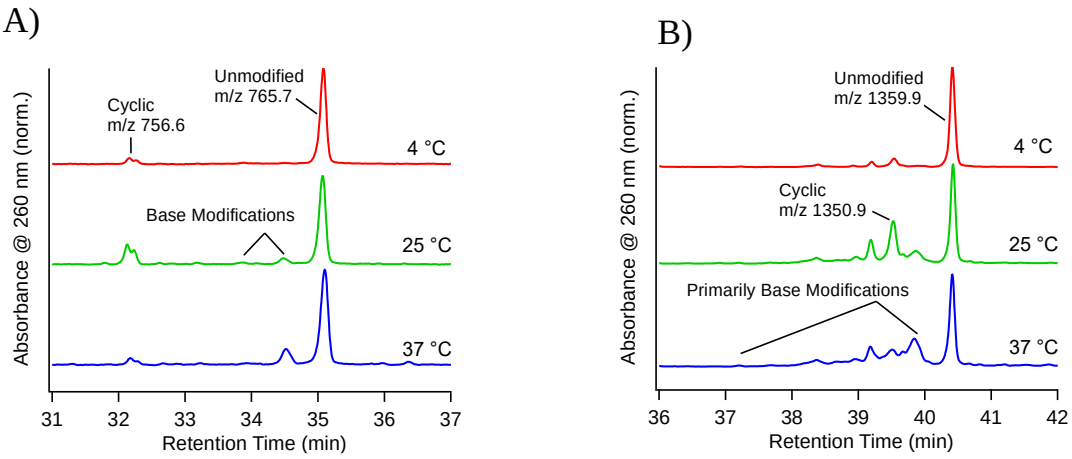

Figure 12. Analysis of substrates modifications by EDC. UV-monitored HPLC chromatograms of A) 5-mer substrate and B) 9-mer substrate after incubation in the reaction buffer for 24 hr at various temperatures. Chromatograms are normalized to the intensity of the main oligomer substrate. The m/z values shown are for the charge number z = 2. The peak identified by a shift in -9 Da is equivalent to a loss of 18 Da typical for formation of cyclic products. Base modifications were determined to be primarily at the N1 of G residues and at the N3 of T residues based on previous reports of EDC reactivity with nucleotides (70).

For all three reaction temperatures the extent of 9-mer modification by EDC at 24 hr was found to be greater than that observed for the 5-mer substrate (Figure 12B). To obtain additional information on 9-mer substrate modification we monitored loss of the unmodified 9-mer substrate as a function of time at each reaction temperature. The initial rate of 9-mer modification, from 0 to 4 hr, increases significantly with temperature (Figure 13A). Such an increase was expected, as the rate of EDC-activated covalent bond formation will increase with temperature and the greater molecular motions at higher temperatures can facilitate the close approach of oligonucleotide ends, which promotes cyclization.

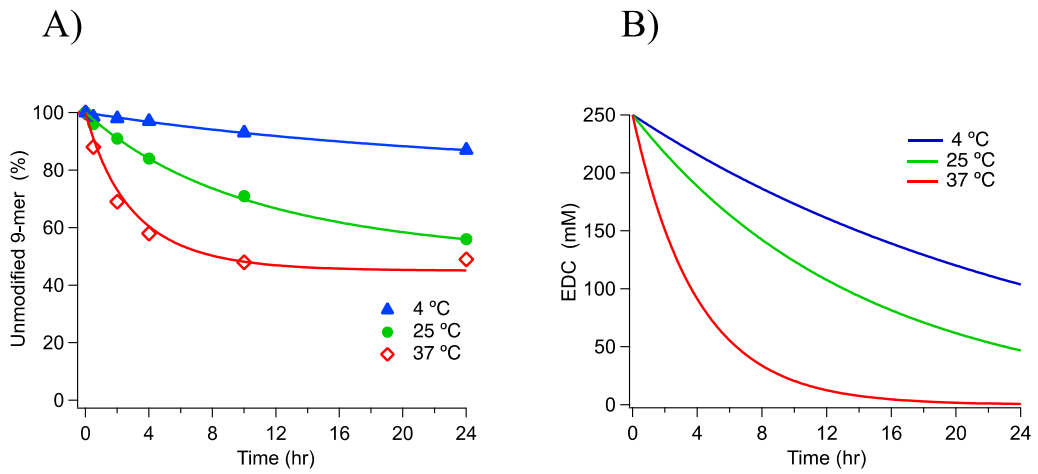

Figure 13. Kinetics of 3′-phosphate 9-mer substrate modification and EDC modification. (A) Results from time course study of remaining unmodified 9-mer substrate as a function of reaction time at the three different temperatures. Percentage of unmodified substrate was found as determined by total integrated intensity of UV-monitored HPLC peaks. (B) Mathematical model prediction of EDC hydrolysis using data provided in A.

At first glance, the data for 9-mer substrate modification appear to exhibit typical exponential decay kinetics for all three temperatures, but this is not the case. At 37 °C modification of the 9-mer substrate ceases by 10 hr, with the same amount (ca. 50%) of unmodified substrate remaining at 24 hr. A simple exponential fit of the 25 °C data indicates that remaining unmodified 9-mer substrate will also reach a minimum of ca. 50%, but not until 48 hr. A similar fit of the 4 °C data indicates that unmodified 9-mer substrate will remain above 75% after several days.

The observed limit to which the 9-mer substrate becomes modified by EDC is likely due to the hydrolysis of EDC over time hydrolysis. As a test of this possible explanation, we modeled the rate of substrate decay while taking into account the temperature dependent EDC hydrolysis rate (details of this model can be seen in section 3.5). Figure 13A provided excellent fits of the data with EDC half-life values of 1.2 days, 15 hr, and 4 hr, and EDC-

concentration dependent side reaction constants of 3.5 x 10^{-5} mM^{-1} hr^{-1}, 2.0 x 10^{-4} mM^{-1} hr^{-1}, and 8.0 x 10^{-4} mM^{-1} hr^{-1} for 4 °C, 25 °C and 37 °C, respectively. A plot of model predicted EDC hydrolysis is shown in Figure 13B. These hydrolysis rates for EDC derived from our 9-mer modification curves are in good agreement with those previously reported (74,75).

The differences between the side product reactivity of the 5-mer and 9-mer substrate can be attributed to both substrate length and base modifications. The shorter the oligonucleotide, the higher the propensity to form cyclized products (15). On the other hand, the longer the oligonucleotide, the higher the probability that it contains bases prone to base modifications.

To gain insight into the relative extent to which substrate modifications by EDC involve the terminal phosphate (e.g., cyclization) versus substrate modifications that do not, oligonucleotides with the sequences of the 5-mer and 9-mer substrates, but without the 3´ phosphate, were analyzed after incubation with EDC. These oligonucleotides are referred to as OH-5-mer and OH-9-mer. Consistent with the observation that the most common modification of the 5-mer and 9-mer substrates at 4 °C is cyclization, OH-5-mer and OH-9-mer exhibited much less modification when incubated with EDC at 4 °C (i.e., <1% for OH-5-mer and about 6% for OH-9-mer after 24 hr) (Figure 14). However, OH-5-mer and OH-9-mer modification was still significant when incubation with EDC was carried out at 37 °C for 24 hr, with OH-5-mer exhibiting 16% modification and OH-9-mer exhibiting 47% modification.

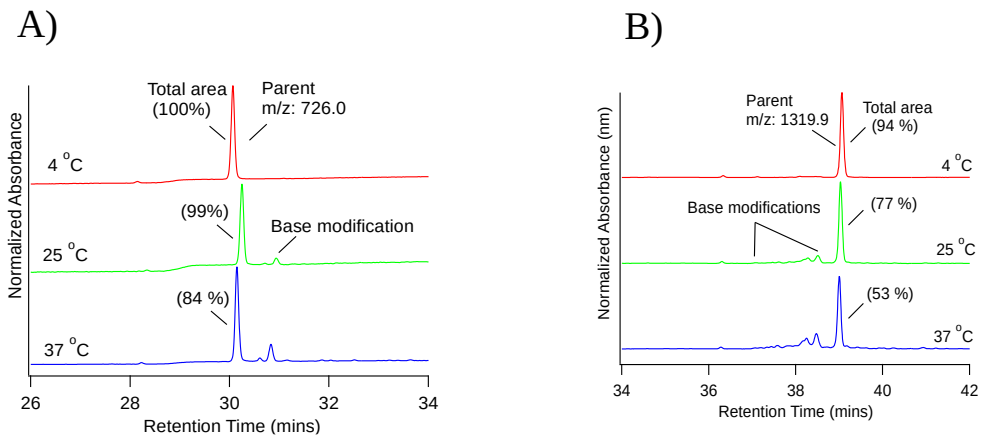

Figure 14. Modifications of the 3´-OH substrate due to EDC monitored using LC-MS. (A) Side product formation for the 3´-OH 5-mer substrate. (B) Side product formation for the 3´-OH 9-mer substrate. All chromatograms are normalized to the intensity of the main oligomer substrate.

The greater extent of OH-9-mer modification compared to OH-5-mer is likely simply due to the greater number of possible sites for nucleobase modification that come with greater oligonucleotide length. However, it is also well known that base modifications by EDC are specific to the G and T bases (27). Thus, we should expect that substrate modification will be strongly dependent on substrate sequence. To illustrate that this is truly the case, control experiments were run with 5-mer oligonucleotides containing only A or C bases, and a third oligonucleotide whose sequence was similar to the 5-mer but modified to have only one base susceptible to modification (Figure 15).

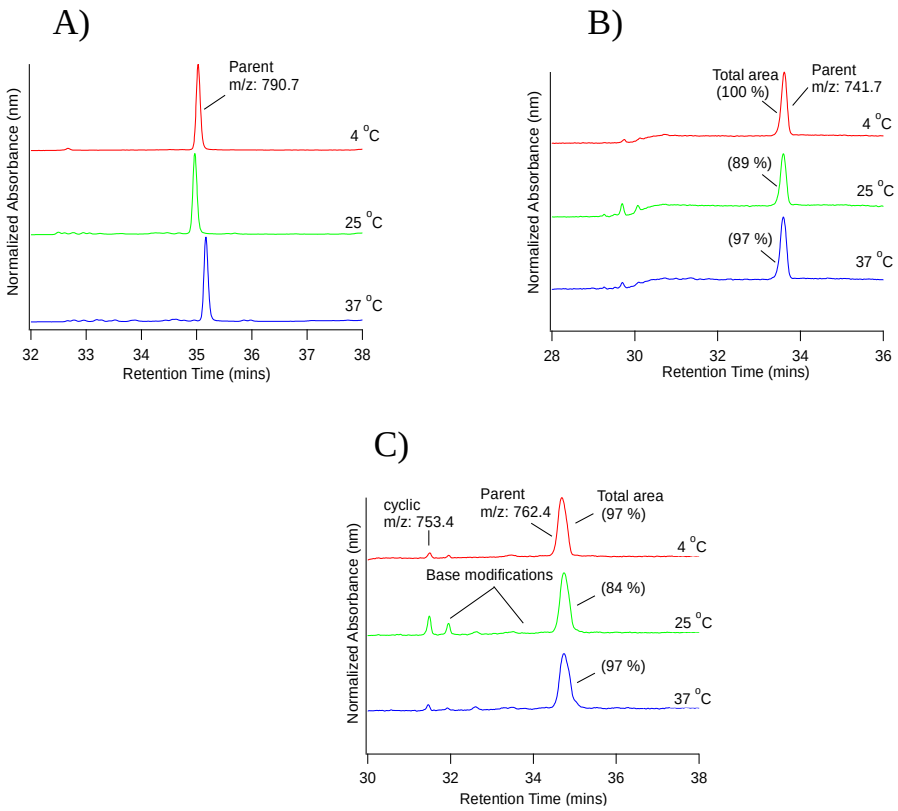

Figure 15. Modifications of the 5-mer substrate due to EDC monitored using LC-MS. (A) Side product formation for the 3´-phosphate A-5 substrate. (B) Side product formation for the 3´-phosphate C-5 substrate. (C) Side product formation for the modified 3´- phosphate 5-mer substrate.

As expected, the removal of G and T bases led to a substantial decrease in the amount of base modifications. In Figure 15A, no modifications can be seen with the A-5 substrate under the reaction conditions. Although base modifications are not expected for the A-5 substrate, the known ability for oligo A to form stacks (76-78) might also have contributed to the absence of cyclization. Figure 15B also shows similarities to Figure 15A, with the exceptions of unmodified modifications ~10% (possibly cyclized substrates) at 25 °C. By engineering the 5-mer substrate so that only one base is susceptible to modification, as in Figure 15C, the amount of unmodified substrate increased from 73 % to 84 % at 25

°C. Altogether, the experiments conducted in Figure 15 show that base modifications can be a significant source of modifications in ligating reaction systems.

While conducting these experiments we noticed that 3´-phosphate oligonucleotides in stock solutions can lose their terminal phosphate group, possibly in association with repeated freezing-thaw cycles of the stock solutions. Dephosphorylation therefore represents yet another mechanism we observed over the course of this study by which substrate modification (or degradation) can lead to reduced ligation yields (Figure 16). The degradation of oligonucleotides due to dephosphorylation can be avoided by storing aliquoted stock solutions at -80 °C to preserve integrity of stock.

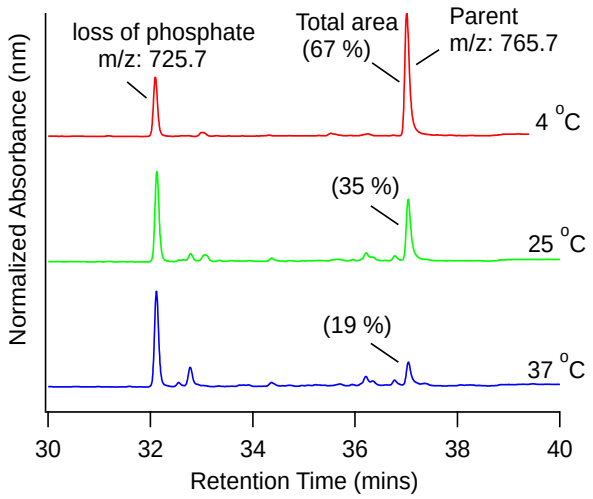

Figure 16. Dephosphorylation of the 3´-phosphate 5-mer substrate.

3.5 Mathematic model of ligation kinetics

Given the effects of equilibrium association and side product formation on ligation yields and kinetics, we sought a mathematical model that could describe the ligation

reaction. By following the reaction scheme in Figure 10, the rate of ligation was modeled using second order kinetics. In the model, the first reaction was assumed to be the hybridization of the substrate and hairpin oligonucleotides (Figure 10A). Afterwards, activation and subsequent ligation of the substrate hairpin complex is modeled (Figure 10C and D). Side reactions of the substrates and hairpin are also included in the model (Figure 10F and G), in addition to hydrolysis of EDC. The model assumes that the hybridization of the substrate and hairpin is fast (Figure 10A), therefore the activation of the substrate by EDC (Figure 10B) is captured through the pathway in Figure 10C. It is also assumed that hydrolysis of an activated phosphate back to a regular phosphate (Figure 10E) does not occur in our reaction system (15).

The following constants used are defined; H represents the concentration hairpin, S is the substrate oligonucleotide, I is the hybridized intermediate complex, P is the ligated products and E is the EDC concentration. The rate constants in the model are forward hybridization rate constant k_1, equilibrium constant K_1, hairpin side product formation rate constant k_2, substrate side product formation rate constant k_3, chemical ligation rate constant k_4, and hydrolysis rate constant k_5.

The overall ligation reaction scheme is defined by the following differential equations:

$$\frac{dH}{dt} = -k_1\left(HS - \frac{I}{K_1}\right) - k_2 HE \qquad (4)$$

$$\frac{dS}{dt} = -k_1(HS - \frac{I}{K_1}) - k_3SE \quad (5)$$

$$\frac{dI}{dt} = -k_1(HS - \frac{I}{K_1}) - k_4IE \quad (6)$$

$$\frac{dP}{dt} = k_4IE \quad (7)$$

$$\frac{dE}{dt} = -k_2HE - k_3SE - k_4IE - k_5E \quad (8)$$

The 9-mer and the 5-mer were fitted separately, based on the information available for each system. For both systems, the equilibrium constant K_1 was obtained as described in Section 2.4. The following steps were then taken to obtain the other rate constants for the 9-mer;

i) The 9-mer side reaction rate constant (k_3) and EDC hydrolysis rate constant (k_5) were obtained by optimizing for the lowest RMSD between the experimental 9-mer decay data (Figure 13A) and using Equations (5) and (8). In the experimental data no hairpin was included therefore all constants relating to the hairpin were ignored during this optimization.

ii) Using the values of k_3 and k_5 obtained above, the chemical ligation rate constant k_4 was then obtained by fitting the 9-mer experimental kinetic data (Figure 4B) with Equations 4-8. The forward hybridization rate constant k_1 was set to $1x10^{10}$ $hr^{-1}M^{-1}$ to ensure the system was in equilibrium, while the hairpin modification rate constant k_2 was made equivalent to k_3 since the stem of the hairpin and the 9-mer substrate have similar number bases exposed to the solvent. k_4 values were found to be 6.5×10^{-4} mM^{-1} hr^{-1}, 0.0021 mM^{-1} hr^{-1}, and 0.0067 mM^{-1} hr^{-1} for 4 °C, 25 °C and 37 °C, respectively.

iii) The value of k_4 was assumed to be equal for both the 5-mer and the 9-mer, since it was assumed that the chemical ligation step will remain the same between these systems.

To fit the 5-mer data the following steps were taken;

i) k_1, k_4, and k_5 were kept constant from the 9-mer optimization

ii) The equilibrium constant K_1 and the side reaction rate constants, k_2 and k_3, for the 5-mer reaction were found by fitting Equations 4-8 with experimental values obtained in Figure 4C. The hairpin side reaction rate constant k_2, and the 5-mer side reaction rate constant k_3, were not assumed to be equal, as in the case of the 9-mer, since the number of exposed bases between the 5-mer and the hairpin stem differed significantly. The values of the optimized constants at 25 °C,

which was used to construct the contour plot shown in Section 3.6 are as follows, K_1 = 2.4 x 10^4 M^{-1}, k_2 = 5.2 x 10^{-5} $mM^{-1}hr^{-1}$, and k_3 = 2.2 x 10^{-6} $mM^{-1}hr^{-1}$.

A summary of rate constants can be seen in Table 4 and Table 5.

Table 4. Summary of rate constants obtained using model prediction for the 9-mer substrate. The hairpin side product rate constant shown here is for hairpin modification when the 9-mer substrate is hybridized.

Reaction Temperature (^oC)	Equilibrium constant, K_1 (M^{-1})	Hairpin side product rate constant, k_2 ($mM^{-1}hr^{-1}$)	Substrate side product rate constant, k_3 ($mM^{-1}hr^{-1}$)	Chemical ligation rate constant, k_4 ($mM^{-1}hr^{-1}$)	EDC hydrolysis rate constant, k_5 (hr^{-1})
4	7.1 x 10^{10}	3.5 x 10^{-5}	3.5 x 10^{-5}	6.5 x 10^{-4}	0.036
25	2.3 x 10^{7}	2.0 x 10^{-4}	2.0 x 10^{-4}	2.1 x 10^{-3}	0.067
37	3.9 x 10^{5}	8.0 x 10^{-4}	8.0 x 10^{-4}	6.7 x 10^{-3}	0.24

Table 5. Summary of rate constants obtained using model prediction for the 5-mer substrate. The hairpin side product rate constant shown here is for hairpin modification when the 5-mer substrate is hybridized. k_4 and k_5 are the same for the 5-mer and 9-mer.

Reaction Temperature (^oC)	Equilibrium constant, K_1 (M^{-1})	Hairpin side product rate constant, k_2 ($mM^{-1}hr^{-1}$)	Substrate side product rate constant, k_3 ($mM^{-1}hr^{-1}$)
4	9.9×10^5	5.0×10^{-5}	1×10^{-6}
25	2.4×10^4	5.1×10^{-5}	2.2×10^{-6}
37	4.0×10^2	1.2×10^{-4}	9.0×10^{-4}

The results of the ligation kinetic fit for the 9-mer and 5-mer are shown in Figure 17. Good agreement can be seen between the model fits and experimental data.

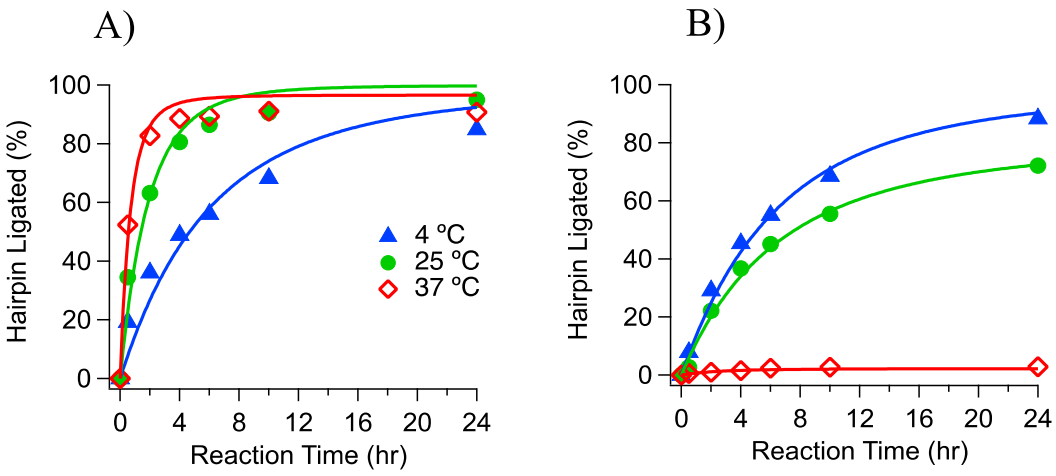

Figure 17. Kinetic plots of the 3′-phosphate systems showing model fits with optimized parameters. (A) Kinetic data for hairpin ligation with the 3′-phosphate 9-mer substrate. (B) Kinetic data for hairpin ligation with the 3′-phosphate 5-mer substrate. Markers are experimental data using a 10:1 substrate hairpin ratio, while solid lines are model fits using optimized rate constants in Table 4 and Table 5.

Based on our experimentally-derived rates for 9-mer substrate modification, EDC hydrolysis, and ligation rates, predicted yields were calculated for 9-mer ligation as a function of substrate:hairpin ratios from 0.25:1 to 2.5:1 for a constant hairpin concentration of 1.3 µM. As can be seen in Figure 18, the resulting prediction of observed ligation yields fit the experimental data for 9-mer ligation at 37 °C.

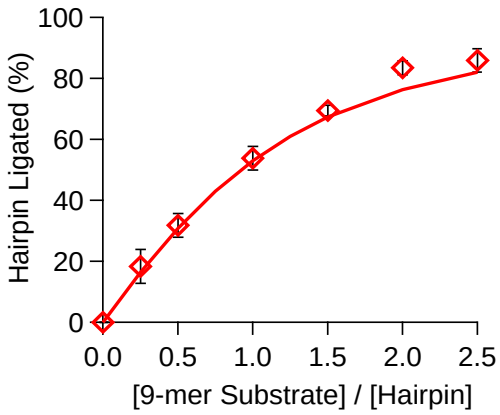

Figure 18. Model prediction of 9-mer ligation at 37 °C as a function of substrate:hairpin ratios. The solid line is the model predicted fit while the markers are experimental data from Figure 5.

3.6 Contour plot: case study for increasing low equilibrium assembly ligation yields

Using the information gained from the mathematical model, 24 hr yields over a wide range of hairpin concentrations and substrate:hairpin ratios for the 5-mer substrate system was predicted.

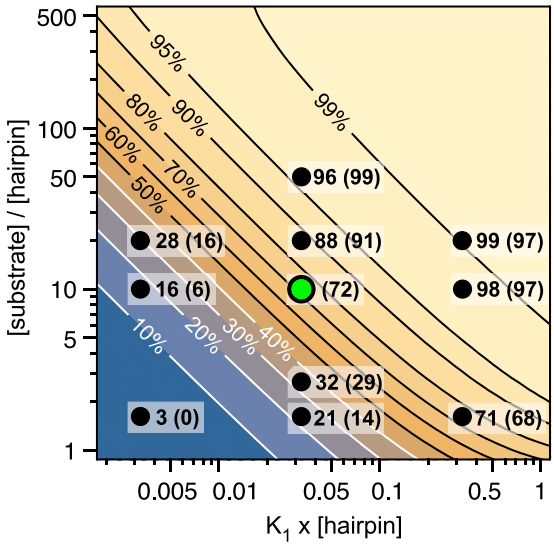

Figure 19. Contour plot showing predicted and experimentally determined ligation yields for the 5-mer 3´-phosphate substrate at 25 ºC after 24 hr in terms of the

dimensionless independent variables equilibrium constant (K_1) x[hairpin] and [substrate]/[hairpin]. The values outside the parentheses are model-predicted yields; values inside parentheses represent experimental yields. The single green point corresponds to the experimental system from which the equilibrium constant and the substrate loss rates were determined.

In Figure 19, we show a contour plot of the predicted ligation yields for the 3´-phosphate 5-mer substrate reaction run at 25 °C for 24 hr as a function of equilibrium constant K_1 and hairpin concentration (their product being the independent variable of the horizontal axis) versus the ratio of substrate and template concentrations (this ratio being the independent variable of the vertical axis).

The mathematical model and RMSD best fits revealed that K_1 of the 5-mer reaction at 25 °C is 2.4 x10^4 M^{-1} , as shown in Table 5. This experimental equilibrium constant data is a single data point on the contour plot of Figure 19, and is shown as a green filled circle (with a horizontal coordinate of 0.031, (K_1 x [hairpin] = 2.4 x10^4 M^{-1} x 1.3 μM), and a vertical coordinate of 10 ([substrate]/[hairpin] = 13 μM/1.3 μM). The other points shown in black filled circles represent both independent experimental yields and predicted yields.

To test the accuracy of our model, we placed points on the contour plot from several 5-mer ligation experiments. For the same horizontal value of K_1 x [hairpin] = 0.031, we plotted points corresponding to 24 hr 5-mer ligation experiments in which substrate:hairpin ratios of 1.5:1, 2.5:1, 20:1 and 50:1 were used (all run at 25 °C, with [hairpin] = 1.3 μM). All four of the experimental yields with K_1 x [hairpin] = 0.031, ranging from 14% to 99%, are within error of their predicted values (Figure 19).

Experiments were also carried out to test the accuracy of this model at different locations along the horizontal axis. Because the plot in Figure 19 is particular to the

experimental conditions upon which the model parameters were derived, it would be difficult to change the equilibrium constant, K_1, without the possibility of also changing other reaction parameters. For example, K_1 could be changed by increasing or decreasing temperature, but a change in temperature would also alter the rate of the chemical steps of ligation. Thus, we changed the horizontal positions of our experiments by increasing and decreasing the hairpin concentration from the original data point (green dot). For these experiments 24 hr ligation yields were measured for hairpin concentrations of 0.13 μM and 13 μM with the 5-mer substrate:hairpin ratios of 1:1, 10:1, and 20:1. At high hairpin concentrations (K_1 x [hairpin] = 0.31) the model predicted yields that were virtually identical to experimental yields. At low hairpin concentrations (K_1 x [hairpin] = 0.0031) the experimental yields were somewhat lower than predicted yields.

It is possible that the predicted lower yields differ more from their corresponding experimental values because experimental errors can be greater when measuring low amounts of product formed, or that the model does not sufficiently capture the impact of substrate and hairpin modifications in low concentration reactions where most of the hairpins are not associated with substrates. Nevertheless, the fact that all the experimental and predicted yields in Figure 19 are within experimental error provides support for our rather simple model being able to predict ligation yields over a wide range of hairpin and substrate concentrations, for a particular system.

To construct a similar predictive model for other ligation systems it would be necessary to determine the kinetic rate constant for bond formation, and the kinetic rate constant for substrate and template loss, for each ligation system. It is possible that the kinetic rate constant for bond formation may be the same for stable complexes using the

same activation chemistry, as appears to be the case for the 9-mer and 5-mer systems studied here. However, the kinetic rate constant for substrate loss could vary tremendously.

3.7 Conclusion

In Chapter 3, a systematic investigation of the combined effects of reaction temperature, template-substrate complex stability, and substrate modification on ligation product yields was conducted. The findings reported in this chapter showed that different reacting systems had different optimum operating reaction conditions. For a low assembling system, as in the case of the 5-mer, a low reaction temperature is required in order to obtain maximum ligation yield. High reaction temperatures in a low assembling system will lead to disassembly of the system and thus perturb the close proximity of the ligation nicks.

Conversely, a high assembling system, as in the case of the 9-mer presents a more complex situation. The reaction temperature can be increased to achieve high ligation rates without concerns of dis-assembling the system. However, high reaction temperatures also lead to the formation of side reactions which will subsequently lead to a loss in the number of substrates available for ligation, especially under equimolar conditions. Additionally, because the 9-mer has four more nucleotides compared to the 5-mer, it is also prone to more significant base modifications which presents yet another avenue to lose a substrate, different from the well-known loss due to cyclization.

Therefore, as discussed in this Chapter, to obtain high yields in any ligating system, finding the optimum reaction temperature is paramount in order to modulate equilibrium assembly, side product formation, and ligation kinetics.

3.8 Materials and Methods

3.8.1 Oligonucleotide sequences

Same as discussed in Section 2.9.1

3.8.2 Ligation experiments

Same as discussed in Section 2.9.4

3.8.3 Side product formation

For monitoring the formation of side products, the substrate oligonucleotides were incubated at 20 μM in the absence of the template hairpin under the regular reaction buffer conditions. At the indicated time point, 10 μL of the reaction mix was immediately zip-tipped using Millipore Sigma C_{18} resin tips to quench the reaction. The decay in the substrate oligonucleotides was then monitored using HPLC and mass spectrometry (MS).

3.8.4 Mass spectrometry

Mass spectrometry was conducted on an Agilent 6430 Triple Quad LC/MS. The mass spectrometer source was set at 300 ºC with an N_2 flow rate of 11 L/min and a capillary voltage of 4000 V. All data was collected in negative ion mode. 10 μL of quenched samples were injected in each run. Separation was done using a Waters BEH-C18 2.1x150 mm column. Column temperature was held at 25 ºC with a flow rate of 0.3 mL/min. Absorbance was measured at 260 nm.

Eluent A is made of 20 mM NH_4HCO_3, 10 mM dibutylamine; Eluent B is made of 20 mM NH_4HCO_3, 10 mM dibutylamine, 33% acetonitrile. 0-10 minutes used isocratic flow at

95% A. 10 – 55 minutes used gradients from 95-24% (A), 95% (A) from 55 – 60 minutes. Products are assigned using expected m/z values.

CHAPTER 4. INTRINSIC NUCLEIC ACID FEATURES INFLUENCE LIGATION YIELDS AND KINETICS

4.1 Introduction

In the previous Chapters, the effects of external factors such as experimental temperature, EDC hydrolysis, and side product formation were investigated to increase ligation yields, with regards to DNA templates and substrates. While researchers interested in optimizing specific DNA ligating systems can benefit from the findings of the previous chapters, it was imperative to understand how broadly these solutions can be applied, particularly in the DNA nanotechnology and origin of life field.

To develop a broadly applicable solution, questions such as which oligonucleotide sequence should be used in the design of the ligating system need to be addressed. Several studies have shown that sequences far from the ligation nick site have little to no impact on ligation yields, compared with the sequences at the ligation nick (30,31). However, there has been no detailed investigation into the effect of nick-site base pairs on ligation, except for the study by Shabarova et al., which showed variations in ligation yields based on the identity of the nick-site base pair for cyanogen bromide ligation systems (42). By focusing on the optimized 9-mer ligation system from previous chapters, this chapter will demonstrate the possible role of nick-site base identity on ligation rates and yields at different reaction temperatures.

The phosphate position on the ligating system is another important factor that will be addressed. In particular, this chapter will investigate the potential benefits and

drawbacks of designing the ligating system with either a 3′p or a 5′p at the ligation nick. A few studies (1,18,(79)) have shown that for DNA system, having a 3′p leads to higher ligation yield than a 5′p, while others have shown that a 5′p leads to better yields (30). None of these studies, however, were performed systematically – investigating a broad range of conditions like temperature which has been shown to yield > 95% product formation in our studies. Although the 5′OH (typically paired with a 3′p for the formation of a 3′-5′ linkage) is a better nucleophile than the 3′OH, by virtue of being a primary alcohol as opposed to a secondary alcohol, the 5′p is more common than the 3′p in biology today. Therefore, in order to understand how non-enzymatic ligation could have occurred on the prebiotic earth, it is imperative that our studies address the possibility of ligation with a 5′p.

The last question that will be addressed in this chapter is the effect of the identity of the sugar on the ligation nick nucleotide. Studies will be designed to investigate the impact of having a ribose, a modified ribose (*2′-O-Methyl* ribose), or a *deoxy* ribose on the ligation rates and yields. Given that each sugar will lead to different reaction intermediates and pathways, understanding the interplay of this relationship in our particular optimized ligation system will be invaluable.

In conclusion, results from this chapter will provide a holistic approach of how intrinsic nucleic acid features such as sequence, phosphate position, and sugars can influence ligation rates and yields.

4.2 Effect of nick-site base pairs on 3′p DNA ligation

For these studies eight hairpins and their complementary 9-mer substrates were used to create nick sites with the four possible flanking GC base pairs (C▼C/GG; C▼G/GC; G▼C/CG and G▼G/CC) and the four possible AT base steps (T▼T/AA; T▼A/AT; A▼T/TA and A▼A/TT). Building on past work in Chapter 2, the 9-mer substrate was chosen at a substrate:hairpin ratio of 1.5:1 to ensure that maximum ligation could be obtained after 24 hr as illustrated by the data presented in Figure 5A. The substrate-hairpin sequences with AT base pairs flanking the nick site were adjusted so that they would have the same %GC content, and therefore similar stability as the set with GC base pairs flanking the nick site, although it is understood that the base pair stability and stacking interactions at the nick site could vary.

The kinetic curves shown in Figure 4B revealed that the 2 hr time point provides a good balance between sufficient product for accurate quantitation of the slowest 9-mer ligation (at 4 °C), while still providing sufficient dynamic range to discern differences in the rates of 9-mer ligation at 25 °C and 37 °C. Thus, for this set of ligation experiments we used the yields measured at the 2 hr reaction time point as a semi-quantitative means for comparing the initial rates of different sequence and the yields measured at the 24 hr time point as an estimate of longer reaction time yields.

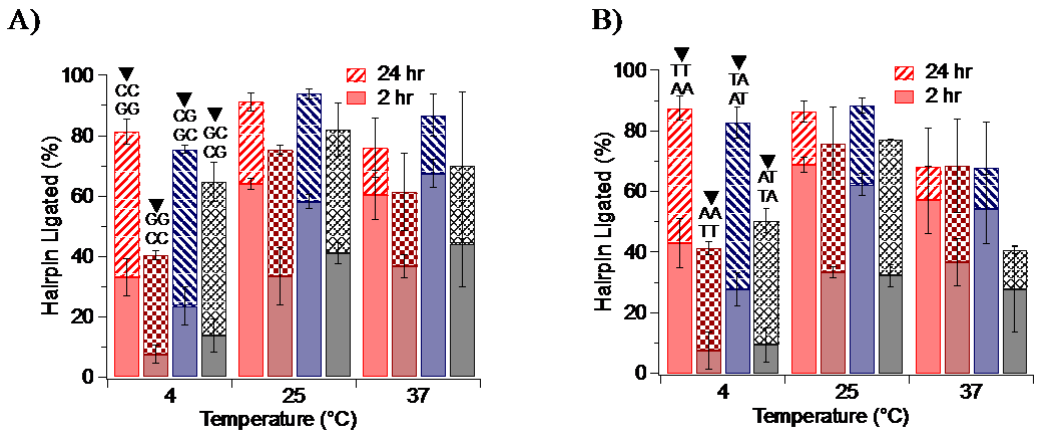

Figure 20. Effects of nick-site flanking base pairs on ligation efficiency (A) Watson-Crick base pairs product formation for the CG substrate:hairpin system at different temperatures. (B) Watson-Crick base pairs product formation for the AT substrate:hairpin system at different temperatures. All reactions were carried out for 2 hr and 24 hr indicated by the unfilled and shaded bars respectively, at 4 °C, 25 °C, and 37 °C, with a substrate:hairpin ratio of 1.5:1 in a buffer containing 5 mM $MnCl_2$, 100 mM MES pH 6.0, and 250 mM EDC. Directionality of sequences is shown in Figure 10H. Each bar graph is from an average of two or three independent experiment replicates; with error bars representing the range of values measured.

The comparison of yields for the 2 and 24 hr reaction time points for the four possible nick sites with flanking GC base pairs reveals a striking difference between the ligation rates of the C▼C/GG and G▼G/CC nick sites (Figure 20A). For example, the initial rate of ligation at a C▼C/GG site is three times the initial rates of a G▼G/CC site at 4 °C and 25 °C. For reactions carried out at 37 °C this difference in the initial rates is decreased but remains significant. Because all substrate:hairpin combinations had similar predicted T_m values, and because observed rates do not correlate with sequence-specific free energies of base stacking at nick cites (80), we expect that the differences observed in initial ligation rates is due to differences in base stacking geometry near the activated phosphate and the alcohol nucleophile.

We speculate that stacking of the G▼G bases across the nick site induces rigidity on the bases and slows ligation by holding the reactive groups away from their optimal reaction geometry, particularly at lower reaction temperatures. This rigidity can be overcome by increasing the temperature of the reaction. Nevertheless, the maximum yield attained for the G▼G/CC nick site is consistently lower than those of the other GC nick sites, even at 37 °C. Additionally, the observed decrease in ligation yields could also be attributed to base modifications of the N_1 substitution of guanine, as these modifications could potentially impact the Watson-Crick pairing of the bases (81).

Similar results are observed for the four possible AT base steps that flank the nick site (Figure 20B). The direct correspondence between GC and AT systems for the impact of pyrimidine-pyrimidine stacking at the flanking site, versus purine-purine stacking, support the hypothesis that these differential initial rates and temperature dependence is due to base stacking, as opposed to non-Watson-Crick hydrogen bonding interactions that would be expected to be different for the GC and AT base steps. These findings are similar to those observed by Carrierio and Damha (44) in which the G▼T/CA nick site ligation site gave 10% yield, while a T▼T/AA nick site gave 66% yield. Other reports have also indicated that pyrimidine bases flanking a nick is better than other base flanking compositions (82,83).

4.3 Mismatches in nick-site base pairs, Tm, and effects on 3′p ligation

Having investigated the Watson-Crick paired system in Section 4.2, we examined if purine-purine and pyrimidine-pyrimidine stacking around a nick site would have a similar impact on ligation at mismatched base pairs.

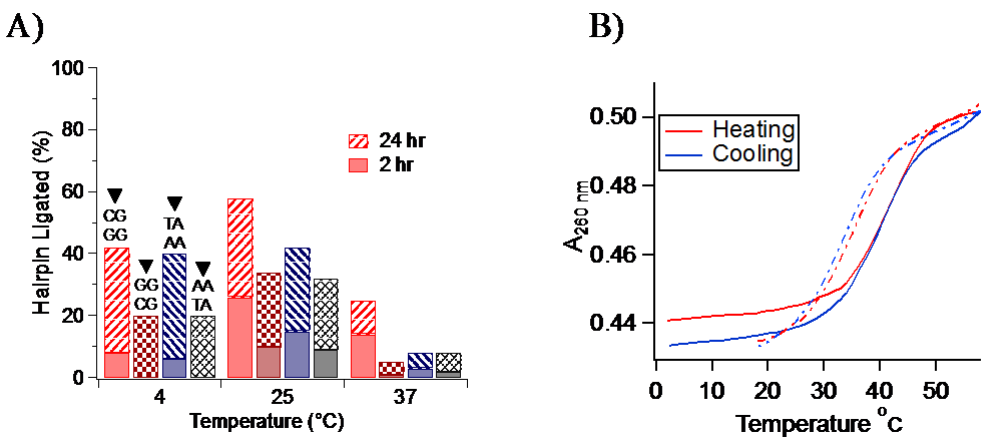

Figure 21. Effects of nick-site flanking base pairs on ligation efficiency and melting temperature (T_m) determination for a substrate:hairpin mismatch system. (A) Mismatch base pairs product formation for both the CG and AT substrate:hairpin system at different temperatures. Reactions were carried out similar to those described in Figure 20. (B) T_m determination of the substrate hairpin system shown in the first bar-graph of Figure 21A. The dashed lines represent the mismatch 9-mer substrate, while the solid lines represent the T_m of the Watson-Crick 9-mer substrate for the same hairpin system. The mismatch 9-mer dissociates from the hairpin at a T_m of 37 °C.

In Figure 21A, it is shown that when a purine-pyrimidine flanking base pair is used, such that the purine stacking is opposite the substrate:hairpin nick, 40% of ligated products are observed at 4 °C. In contrast, when a purine-purine flanking base pair is used with a purine mismatch, the yield is decreased by half to 20% at the same temperature. As observed previously, an increase in reaction temperature to 25 °C led to an increase in initial rates and maximum yields. However, at 37 °C, the incorporation of mismatches is significantly reduced to less than 10% in most cases. This drastic decrease in reactivity at 37 °C can be attributed to the lower melting temperature the mismatch nucleotide introduces to the system (i.e. the T_m of the main substrate:hairpin sequence decreases from 45 °C to 37 °C) (Figure 21B).

It is important to note that the incorporation of mismatched base pairs at a nick site (Figure 21A) can occur with an initial ligation rate and maximum yields that are comparable to that of fully Watson-Crick duplexes that have two purine bases flanking the nick site (Figure 20A and B). These findings corroborate studies which have previously found that chemical ligation is quite tolerant to the introduction of mismatches (73,84).

4.4 Influence of phosphate position on DNA ligation

4.4.1 Difference in reaction mechanism

All the experiments in this Chapter, so far, have been conducted with a 3′p, in this section, we will investigate the effect of changing the ligating system to a 5′p. In order to obtain comparable results to those of the 3′p, the hairpin template will be phosphorylated on the 5′ end using T4-PNK (T4 Polynucleotide Kinase). This phosphorylated hairpin template will then be ligated to a 9-mer substrate with a 3′-OH end.

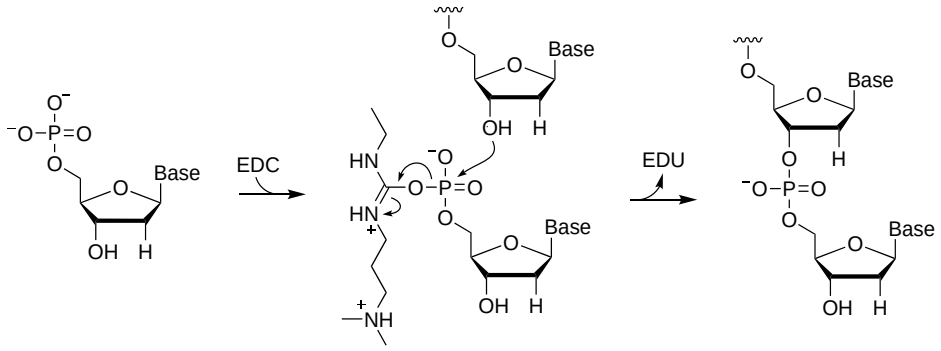

Figure 22. Key steps in the reaction mechanism for 5´-phosphate substrate activation by EDC and subsequent ligation.

The ligation mechanism for a 5′p ligation is demonstrated in Figure 22. The 3′-OH nucleophile attacks the activated phosphate after incubation with EDC. A major drawback of this system is that the 3′-OH nucleophile is known to be a weaker nucleophile compared to the 5′-OH (15) and could therefore lead to the formation of fewer ligation products. However, an advantage of this system design is that due to the length of the hairpin template cyclization is less likely to become a major side product of the reaction, unlike the 3′p system, discussed in section 3.4.

4.4.2 Ligation kinetics and influence of nick-site base pairs

Kinetic experiments were conducted to monitor how the ligation yield varied over time. In Figure 23A, it is observed that an increase in reaction temperature from 4 °C to 37 °C led to an increase in the instantaneous reaction rate of the ligation reaction. Similar to section 2.5, it is observed that the reaction at 37 °C reaches a plateau quickly compared to the reactions at 4 °C and 25 °C, most likely due to the higher rate of formation of side product. We note that although cyclization is not the main side product result in this system, base modifications will still have a huge impact on both the hairpin template and the substrate. A major finding from this study that under all the reaction conditions studied, the maximum ligation yield for this system is only 40%. This is a large decrease from the 95% maximum yield observed for the 3′p system. It is possible that the ligation yields might improve if the reaction was left for longer than 48 hr, however, the half-life of EDC renders this unlikely.

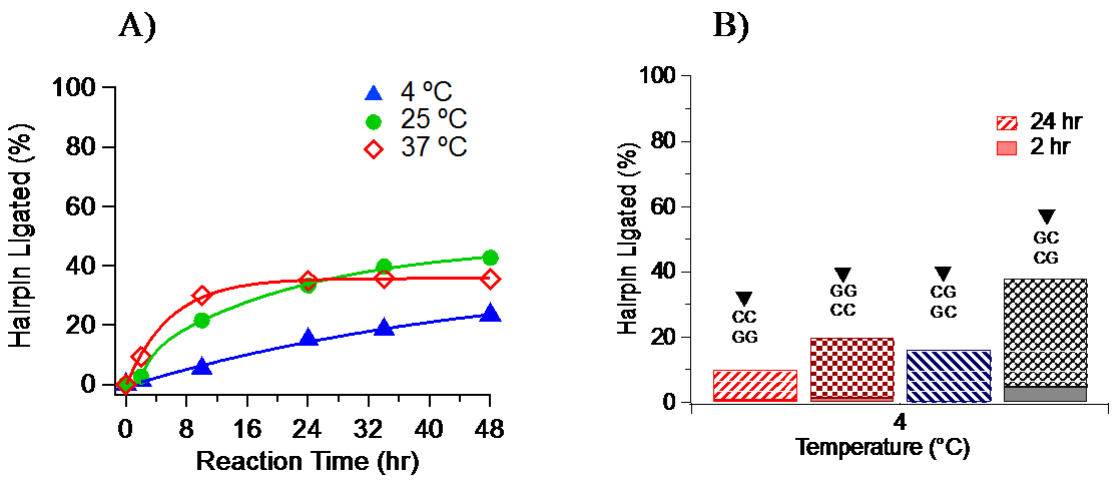

Figure 23. Kinetic measurements of chemical ligation at different temperatures and nick-site reactivity. (A) Kinetic data for hairpin ligation with the 3´-hydroxyl 9-mer substrate and 5´-phosphate hairpin template. Buffer conditions: 5 mM $MnCl_2$, 100 mM MES, pH 6.0 and 250 mM EDC. 1.3 µM hairpin and 13 µM of substrate DNA oligonucleotides were used. (B) Watson-Crick base pairs product formation for the CG substrate:hairpin system at 4 °C. Buffer conditions are the same as in (A) except with a substrate:hairpin ratio of 1.5:1.

Despite the low reaction yields, we sought to identify whether we could observe any selectivity in nick-site base pairs reactivity. Contrary to section 4.2, in Figure 23B, we do not observe a stark difference in the reactivity of the C▼C/GG and G▼G/CC nick sites. There is no difference between having purine stacks or pyrimidine stacks at the ligation nick, the reason for this is not fully understood. However, there could be a difference in reaction site geometries for the two phosphate positions which is not captured by our current experimental analysis.

4.5 Effect of the 2′ terminal molecule on 3′p and 5′p ligation

In order to demonstrate the effects of our studies on other types of nucleic acid systems, we designed studies that investigated the differences in ligation yield when the 2′ terminal molecule was changed on the 3′p system. Our optimized system design of a

hairpin template and 9-mer substrate was used. Three types of 9-mer substrates were studied, the first type in which all the substrates were DNA and has been used previously in all our studies. The second type of substrate was similar to the first in that all the nucleotides are DNA except the last nucleotide, which was RNA, this substrate is herein referred to as rDNA. The last type of substrate was similar to the rDNA substrate, except with the addition of a 2′-O-Methyl to the 2′ end, this substrate is herein referred to as 2′-O-Me rDNA.

We sought to understand how the different substrates, which would potentially have different chemistries could affect the ligation yield. HPLC traces were used to identify the reaction intermediates after the substrates were incubated with EDC. In Figure 24A and C, it can be observed that for the DNA and 2′-O-Me rDNA substrate, there is no stable intermediate formed upon incubation of the substrates with EDC. The absence of a stable intermediate was further verified by the co-injection of both the pure substrate and the substrate incubated with EDC which resulted in only one peak.

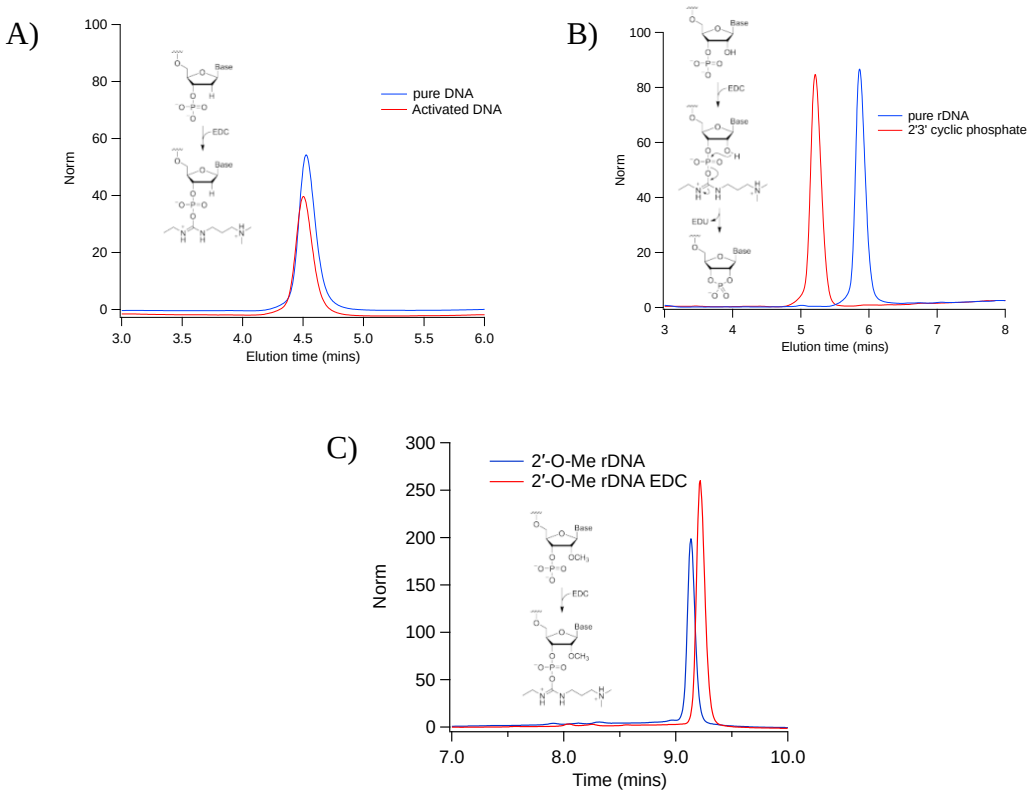

Figure 24. HPLC trace of the (A) DNA, (B) rDNA, and (C) 2′-O-Me rDNA substrate oligonucleotide incubated without the hairpin template in the presence of EDC at 4ºC for up to 12 hours. The inset in each plot shows the expected reaction chemistry for each substrate type.

On the contrary, in Figure 24B the formation of a different peak can be observed upon addition of EDC. This new peak was confirmed to be present after co-injection of the substrates with and without EDC and was identified by mass spectrometry as a cyclic phosphate. Unlike the DNA and 2′-O-Me rDNA substrate, the formation of a cyclic phosphate is prevalent in the rDNA substrate because an activated phosphate is more prone to intramolecular attack by the nearby 2′ OH commonly present in RNA nucleotides. Given previous studies which have shown that a cyclic phosphate could be prebiotically relevant, we proceeded to investigate whether the rDNA substrate will lead to higher ligation results for our ligating system (22, 41,42).

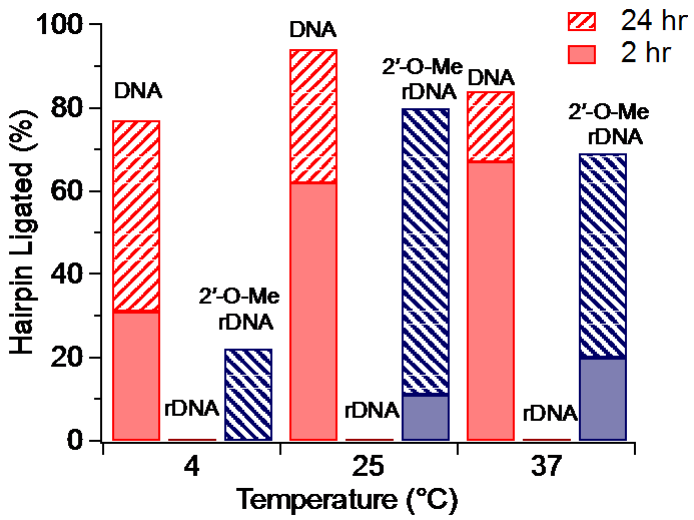

Figure 25. Effects of the 2′ terminal molecule on ligation efficiency for a 3′-phosphate system. Reactions were carried out for 2 hr and 24 hr indicated by the unfilled and shaded bars respectively, at 4 °C, 25 °C, and 37 °C, with a substrate:hairpin ratio of 1.5:1 in buffer containing 5 mM $MnCl_2$, 100 mM MES pH 6.0, and 250 mM EDC.

In Figure 25, it is observed that at 4 °C the DNA oligonucleotide has a higher instantaneous rate (approximated by the yield after 2 hr) and forms more products compared to 2′-O-Me rDNA and rDNA substrates. At higher temperatures, the 2′-O-Me rDNA and the DNA substrate approach the same maximum despite the fact that the instantaneous rate of the DNA oligonucleotide remains higher. These results caution against using the instantaneous rate of ligation as the sole indicator for determining ligation efficiency. The increased reactivity of the 2′-O-Me rDNA at elevated temperatures further demonstrates that higher temperatures can allow the activated oligonucleotides to attain more favorable conformations which can lead to ligation. The differences in the rates of ligation and product formation observed in Figure 25 is attributed mainly to the functional groups since the theoretical Tm's of the different oligonucleotides are similar (within ± 2 %).

Surprisingly, despite cyclic phosphate formation, the rDNA substrate does not lead to a significant amount of product formation under these reaction conditions. Rather than act as an activated intermediate for ligation, this cyclic phosphate intermediate appears to act as an inhibitor which hinders the ligation reaction. The yield of the rDNA oligonucleotide can also be improved from <5% to 20% by increasing the pH of the ligation reaction from pH 5.5 to pH 10, as demonstrated in Figure 26. Nevertheless, the formation of only 20% ligated product by the rDNA oligonucleotide demonstrates poor reactivity for rDNA compared to DNA and 2′-O-Me rDNA substrates. Altogether, our results demonstrate that the 2′,3′ cyclic phosphate is not an efficient activated intermediate for ligation reactions under robust conditions. A point that will be further investigated in Chapter 5.

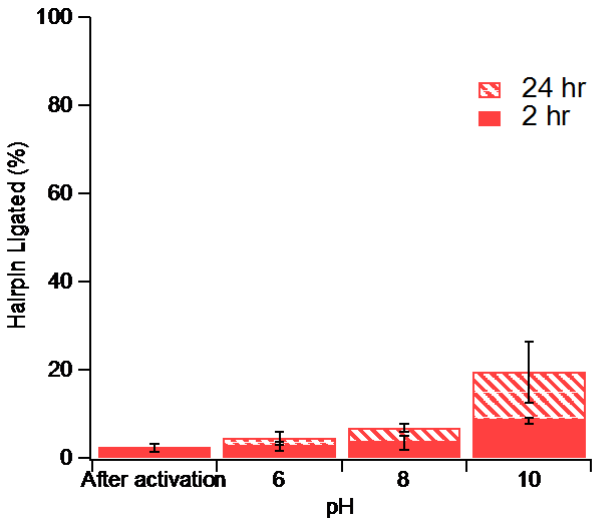

Figure 26. Effect of an increase in reaction pH on rDNA substrate ligation reaction. Experiments were conducted at 25 °C in a buffer containing 5 mM $MnCl_2$, 100 mM MES pH 6.0, and 250 mM EDC originally. After 6 hr of reaction time, a 10x fold dilution was made to increase the pH of experiment. The error bars represent the standard deviation of three independent experiment replicates.

We hypothesized that the absence of the 2′,3′ cyclic phosphate on the oligonucleotides could lead to an increase in rDNA ligated products. Therefore, using the same sequences as in our previous results we investigated ligation reactions involving a 5′p templates and 3′ OH substrates.

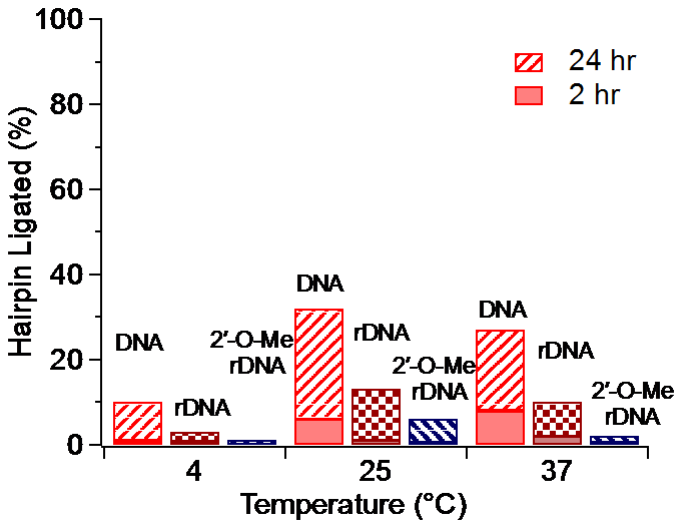

Figure 27. Effects of the 2′ terminal molecule on ligation efficiency for a 5′-phosphate system. Reactions were carried out for 2 hr and 24 hr indicated by the unfilled and shaded bars respectively, at 4 °C, 25 °C, and 37 °C, with a substrate:hairpin ratio of 1.5:1 in buffer containing 5 mM $MnCl_2$, 100 mM MES pH 6.0, and 250 mM EDC.

In Figure 27, we observe a general decrease in the formation of products for DNA complementary oligonucleotides in the 5′p system compared to the 3′p system. The maximum products obtained overall, after 24 h is only 30%, compared to 95% products observed for the 3´p system. There is also a significant decrease in the products formed for the 2′-O-Me rDNA (i.e. the products decrease from 80% in the 3´p system to 10% in the 5´p system). Remarkably, only the rDNA substrate shows an increase in product formation in the 5′p system. Particularly, the rDNA oligonucleotide forms 15% of the ligated products at 25 ^{o}C without inducing high pH conditions needed to obtain similar ligation products in

the 3´p system. Evidently, the inability to form the cyclic phosphate led to the subsequent increase in the formation of ligated products.

4.6 Kinetics of rDNA and RNA ligation

Building on our success of obtaining ligation for 5′p and 3′ OH rDNA system, we were interested in observing how these yields could change over time. The rDNA substrate was incubated with a DNA hairpin template, while the RNA substrate was incubated with an RNA hairpin template.

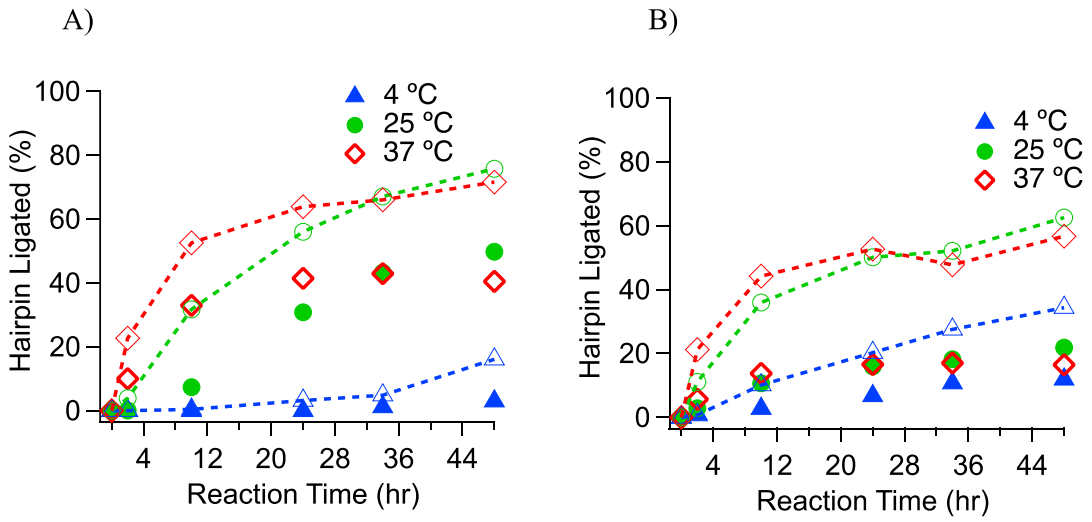

Figure 28. Kinetic measurements of chemical ligation at different temperatures for the rDNA and RNA substrates. (A) Kinetic data for ligation with the 3´-phosphate rDNA 9-mer substrate and 5´-phosphate DNA hairpin template. (B) Kinetic data for ligation with the 3´-phosphate RNA 9-mer substrate and 5´-phosphate RNA hairpin template. Markers (bold-faced) indicate experimental results for reactions without the addition of 150 mM 1-Ethyl, while open faced markers traced by dashed lines indicate results with the addition of 1-Ethyl. Buffer conditions: 5 mM $MnCl_2$, 100 mM MES, pH 6.0 and 250 mM EDC. 1.3 µM hairpin and 13 µM of substrate oligonucleotides were used.

In Figure 28, it can be observed that a general increase in reaction temperature leads to an increase in ligated products, as has been observed previously. Without the addition

of 1-Ethyl (bold-faced markers), the reaction is slow, especially at 4 °C where the maximum reaction yield after 48 hr is 15% for the rDNA and less than 10% for the RNA system. For the rDNA system, Figure 28B, increasing the reaction temperature doesn't lead to a large increase in ligated products, the maximum ligated product at all temperatures hover around 20%. On the contrary, in the RNA system, Figure 28A, increasing the reaction temperature from 4 °C to 37 °C drives the ligation yield up from less than 10% to 40%.

The addition of 1-Ethyl (open faced markers connected by dashed lines) to both systems, demonstrates an increase in ligation products. The maximum yield of the rDNA system increases from 20% to 60%, while that of the RNA system increases from 40% to 80%. This two to three-fold increase in ligation yield is contrary to what was observed for the DNA system (see section 2.7.1), but in line with other studies on RNA ligation (19,21).

The stark difference between the RNA and rDNA system cannot fully be explained. One possibility is that the orientation of the activated phosphate might be in a more favorable conformation in the RNA system, compared to the rDNA. Another possibility, which was not explored in this chapter could be the difference in the backbone conformation. The RNA backbone is known to be in the A-form (85,86), whereas the DNA backbone under aqueous and low salt conditions is known to be mostly in the B-form, (87,88). Therefore, it could be expected that the activated phosphates of rDNA and RNA substrates will attain different conformations due to the differences in the rigidity of the backbone.

4.7 Conclusion

In this Chapter, we investigated the role of intrinsic nucleic acid features such as nick-site base pairs, phosphate position, and sugars on ligation rates and yields. Our studies demonstrated that for the DNA 3´p system, a pyrimidine-pyrimidine step at a nick site (C▼C or T▼T) leads to higher ligation rates than a purine-purine step (G▼G or A▼A), with purine/pyrimidine and pyrimidine-purine steps exhibiting intermediate rates. We note that these trends were not transferrable to the DNA 5´p system, possibly due to different conformations of the activated phosphates.

With regards to phosphate positions, the studies presented here show, quantitatively, the dramatic difference in yields based on phosphate location at a nick site. Consistent with a previous report on DNA chemical ligation (42), our results also confirm that having the terminal phosphate on the 3´ end of one of the two oligonucleotides to be ligated is beneficial. We observe that the rates of bond formation can result in a difference in maximum yield being 95%, in the case of a 3´ phosphate, versus 40% for precisely the same DNA system, except for the placement of the phosphate on the 5´ terminus of the nick site.

Lastly, we investigated the reactivity of the nucleotide at the nick of ligation. Particularly, we tested whether the formation of the cyclic phosphate intermediate led to an increase in product formation. Two factors were observed that caused a decrease in the formation of the ligated products, an increase in the bulkiness of the molecule on the 2´end and the formation of the stable cyclic phosphate intermediate (i.e. on the 2´ functional molecule, H >> OCH_3 >> OH). The 2′-O-Me rDNA oligonucleotide is hypothesized to have a slower instantaneous ligation rate compared to the DNA substrate because the presence of the methyl group interferes with the conformation of the hybridized duplex

during ligation. Likewise, the rDNA substrate is a worse substrate compared to the DNA and 2′-O-Me rDNA because of the presence of a bulky 2´OH, and due to the formation of a stable cyclic phosphate which hinders the formation of ligation products. We observed an increase in product formation for the rDNA and RNA substrate when a 5´p was used due to its inability to form the stable 2´,3´ cyclic phosphate intermediate. These results suggest that ligation with the use of 2´3´ cyclic phosphate intermediate is only favorable under high pH conditions (51), or in evolved systems such as the Hammerhead ribozyme and viruses (25,89), a point that will be investigated further in Chapter 5.

In summary, the results presented in this Chapter demonstrate the applicability of the ligation principles developed in Chapters 2 &3 to other types of nucleic acid systems.

4.8 Materials and Methods

4.8.1 Oligonucleotide sequences

DNA 3´p system

CG Hairpin:	5´ CAGTCACGGAACGTGACTGGACAGGAGA 3´ 6-FAM
GC Hairpin:	5´ GAGTCACGGAACGTGACTCCACAGGAGA 3´ 6-FAM
GG Hairpin:	5´ GAGTCACGGAACGTGACTCGACAGGAGA 3´ 6-FAM

CC Hairpin:	5´ CAGTCACGGAACGTGACTGCACAGGAGA 3´ 6-FAM
AT Hairpin:	5´ ACGTCACGGAACGTGACGTTACAGGAGA 3´ 6-FAM
TA Hairpin:	5´ TCGTCACGGAACGTGACGAAACAGGAGA 3´ 6-FAM
TT Hairpin:	5´ TCGTCACGGAACGTGACGATACAGGAGA 3´ 6-FAM
AA Hairpin:	5´ ACGTCACGGAACGTGACGTAACAGGAGA 3´ 6-FAM

The sequences highlighted in orange are in the loop of the hairpin, while those highlighted in blue showcase the differences in the hairpins. For the naming of the hairpin, the first letter represents the base at the 5´-OH terminal, while the second letter represents the pairing partner for the substrate. For example, in the GC Hairpin, G represents the sequence at the 5´-OH terminal, while C represents the pairing partner for the substrate. Depending on whether a Watson-Crick base-pair or a mismatch was needed, the following substrates were used.

C 9-mer: 5´ TCT CCT GTCp 3´

G 9-mer: 5´ TCT CCT GTGp 3´

A 9-mer: 5´ TCT CCT GTAp 3´

T 9-mer: 5´ TCT CCT GTTp 3´

DNA 5´p system

For the 5´p system, the hairpin templates described earlier were phosphorylated by using T4 Polynucleotide Kinase (New England Biolabs). 9-mer substrates were purchased from IDT without the 3´- phosphorylation.

rDNA and 2′-O-Me rDNA substrates

The methods described above to obtain phosphorylation and dephosphorylation were also implored in these systems. The sequences are identical to the C 9-mer and described below,

rDNA 9 mer: 5´ TCT CCT GT**rC**p 3´

The bolded sequence indicate which nucleotide was a ribonucleotide. The other sequences were deoxyribonucleotides.

2´-O-Me rDNA 9 mer: 5´ TCT CCT GT**rC**p 3´

The bolded sequence indicate which nucleotide was a 2´O methyl nucleotide. The other sequences were deoxyribonucleotides.

RNA 3´p and 5´p system

The sequences for the RNA system were identical to those for the DNA and purchased directly from Integrated DNA Technologies.

4.8.2 Ligation experiments

Same as discussed in Section 2.9.4

4.8.3 HPLC Analysis

All HPLC analysis was conducted on an Agilent 1260 Infinity HPLC system. A DNA PAC^{TM} PA 200 4 µm Anion Exchange Column (4x250 mm) was used for analysis. The following buffers and conditions were used for all analyses: (A) 12.5 mM Tris, pH 8.0; (B) 12.5 mM Tris, pH 8.0, 1.5M NaCl. Flow rate 1.20 mL/min, 10 µL sample injection, and column at 25 ^{o}C. Absorbance was measured at 260 nm and calculation of peak areas were carried out using Chemstation B.04.03.

The following elution scheme was used: Isocratic run of 95% (A) for 5 minutes, gradient from 85 to 55% (A) from 5 to 20 minutes, 95% (A) from 20 to 24 minutes. Peak assignments were conducted by running oligo standards and confirming new peaks through spiking and MS analysis (see section 3.8.4 for Mass spectrometry discussion).

CHAPTER 5. INVESTIGATING THE ROBUSTNESS OF CYCLIC PHOSPHATE LIGATION

5.1 Introduction

The focus of this Thesis Chapter is to investigate the robustness of cyclic phosphate ligation, with particular emphasis on its relevance to the prebiotic world. Recombination of RNA is thought to be among the earliest mechanisms for selection of structure and function according to some proponents of the RNA world (83,90-92). In particular, this reaction proceeds through a near-energy neutral transesterification pathway involving spontaneous cleavage to form a 2′-3′ cyclic phosphate and subsequent ligation of the RNA strand (23,93,94). In addition to ligation reactions, cyclic phosphate intermediates have also been used in nucleotide synthesis pathways (22,28-31), in extension reactions (95), and more recently, for the proposed assembly of ribozymes (33). The use of cyclic phosphates in ligation is perhaps inspired by ribozymes such as the hairpin ribozymes and hammerhead ribozymes which have been shown to induce self-ligation after cleavage, often under eutectic ice conditions (25,89,95-98). However, the ubiquity of cyclic phosphates is not limited to prebiotic synthesis and replication of RNA, as naturally occurring processes such as tRNA splicing use similar ligation techniques (99-102).

As highlighted above, extensive work has already been done in the area of cyclic phosphate ligation. However, progress in this field has been stalled by the low yields (often less than 10%) that occur in non-ribozyme systems (24,51,82). These low yields have previously been attributed to cyclic phosphate hydrolysis (23,83,93,103,104), inability of

the substrates to attain the right conformation for ligation, especially in a duplex format (24,51,105), and the prevalence for the formation of 2´-5´ linkages which are more prone to hydrolysis over the canonical 3´-5´ linkages (27,106).

Notably, work from the Vlassov group (24,27,50,51,107) over the past decade has focused on understanding the efficacy of spontaneous cleavage and ligation of RNA substrates in a template directed fashion. A major finding from their work was that adjacent end-end ligation of substrates does not often result in a favorable conformation for ligation, instead the formation of loops at the ligation site were more common (24,51). Subsequent studies (82,83) have also highlighted the appearance of loops in cleavage and ligation systems. In particular, work by Mutschler and coworkers (82) found that up to a 10-fold enhancement in ligation rate was observed for a particular loop ligation system when compared to its end-end ligation counterpart, under the same reaction conditions.

Intrigued by the finding that ligation in loops is prevalent in cyclic phosphate ligation and given the perceived role of cyclic phosphates in the prebiotic world, we designed a systematic study to investigate the role of sequences and loops in recombination reactions and to understand the robustness of cyclic phosphate ligation. In particular, the 5´-UAA-3´/5´-GAA-3´ loop was chosen as the framework for describing our ligation reactions, inspired by its prevalence in the ribosomal RNA (108-110) and the possibility that non-enzymatic recombination of loops could have played a role in the assembly of the ribosome (111,112).

5.2 Design of a template and substrate system

The standard ligation system was designed to study different cyclic phosphate positions in the ribosomal loop, as shown in Figure 29A. All reactions consisted of an RNA template and two RNA substrates as shown in Figure 29B. The left-handed substrate is the cyclic phosphate substrate and is ligated to the right-handed substrate which has a 5´-OH terminus and a fluorescein label on the 3´ terminus. Given the use of Mg^{2+} for ligation (27), it is expected that some cyclic phosphate substrate will be lost to hydrolysis (Figure 29C) prior to ligation (Figure 29D).

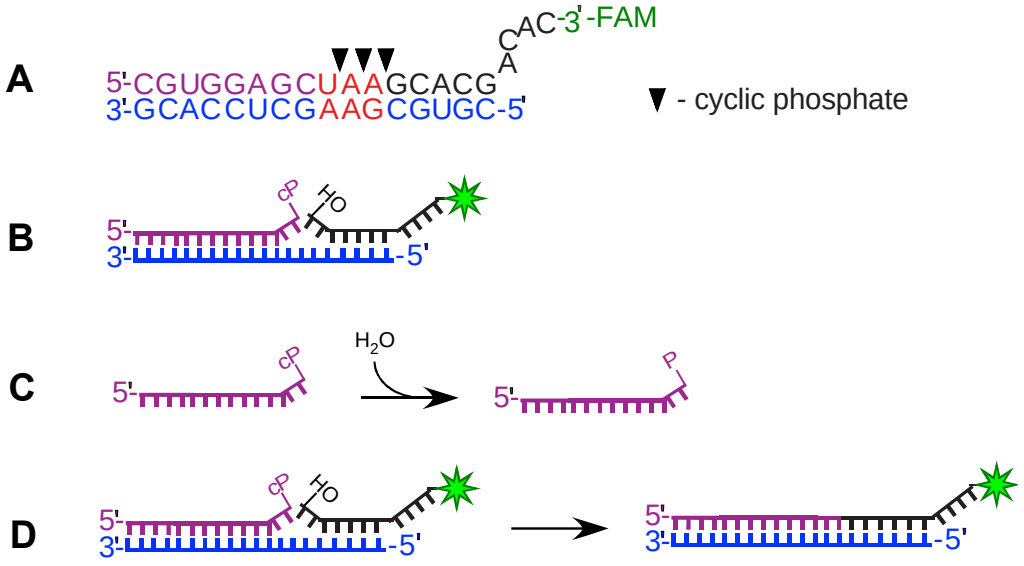

Figure 29. Key reactions in the RNA ligation system. (A) Sequence of templates and substrates used for the ribosomal ligation. The black triangles represent the three cyclic phosphate position investigated for the ribosomal system. (B) Template and substrates used in all ligation systems. (C) Hydrolysis of the cyclic phosphate substrate. (D) Ligation of the substrates to form a product.

For all reactions, the cyclic phosphate substrate was first generated using EDC, as described in Section 5.10.1, prior to its use in ligation, this mechanism is shown in Figure 30. Except otherwise indicated, the FAM labeled substrate and the template were present at 1 μM, along with 2 μM of the cyclic phosphate substrate, based on preliminary studies

that showed that ligation using a 2:1 molar ratio of cyclic phosphate substrate to template would produce a good comparison of yields for most ligating systems (Figure 31). After incubation in the appropriate buffer, the reaction mix was then moved to one of two incubation temperatures, 4 ºC or 25 ºC. All experimental yields were characterized using gel electrophoresis with a sample gel shown in Figure 31A.

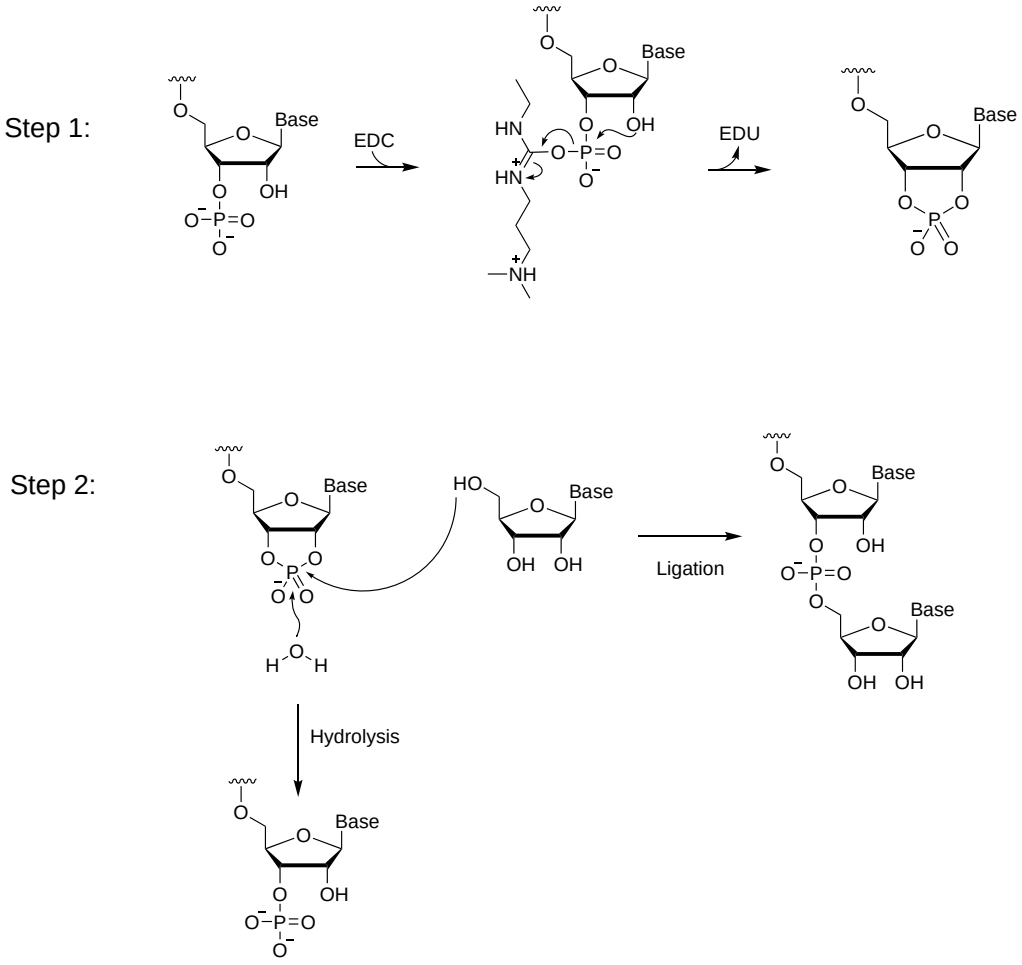

Figure 30. Two-step reaction scheme showing key reactions in ligation system. Note the absence of a template in scheme 2 does not reflect experimental conditions.

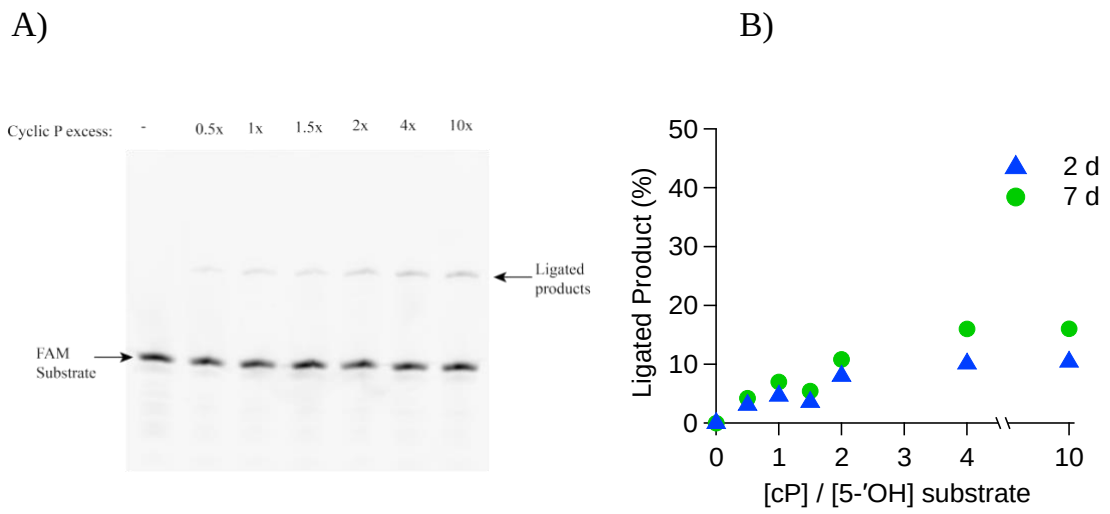

Figure 31. Ligation of a 5´-UUpA-3´/5´-UAA-3´ system. (A) Gel electrophoresis image of the ligation reaction at 4 ºC (B) Reaction kinetics at 4 ºC. The 5´-OH FAM substrate and the template were held constant at 1 µM in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. Data points were obtained by scanning of polyacrylamide gels such as the examples shown in Panel A.

5.3 Ligation at different nucleotides in ribosomal loops

In the 5´-UAA-3´/5´-GAA-3´ ribosomal loop, it was not clear a priori where cleavage would occur in order to reveal the cyclic phosphate in a ribosomal RNA. Therefore, experiments were designed to study cyclic phosphate ligation at different points within the loop. In particular, ligation was studied between the UpA, ApA, and ApG junction (the G junction is shown in Figure 29A) as demonstrated in Figure 32.

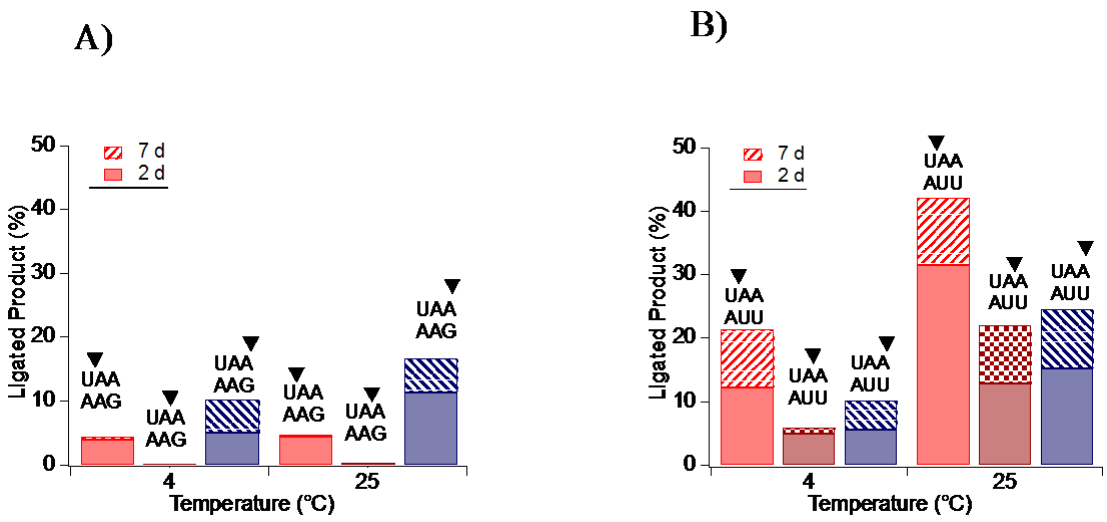

Figure 32. Ligation yields due to cleavage at different nucleotides in a ribosomal loop compared to the Watson-Crick base pair system. (A) The ribosomal GAA system. (B) The Watson-Crick UUA system. All reactions were carried out for 2 days and 7 days indicated by the unfilled and shaded bars respectively, at 4 °C, and 25 °C. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. The FAM substrate and the template were in equimolar amounts of 1 µM. The directionality of the labeled substrates is 5´-UAA-3´ (top label) /5´-GAA-3´ or 5´-UUA-3´ (bottom label). The black triangles represent the cyclic phosphate position.

The maximum amount of ligation that is observed for the ribosomal 5´-UAA-3´/5´-GAA-3´ system is 18%, when the ligation nick is an ApG (Figure 32A). There is no appreciable difference between yields at 4 ºC and 25 ºC for the 5´-UAA-3´/5´-GAA-3´ system. The location of the ligation nick between two A's (ApA) does not lead to ligation. Lastly, no difference in yield is observed between the yields for a UpA system at 2 d and 7 d for both reaction temperatures studied, with ligation yields remaining at 5% under all four conditions.

In Figure 32B, for the Watson-Crick system 5´-UAA-3´/5´-UUA-3´, higher ligation yields overall are observed in comparison to the ribosomal system. The lowest yielding system occurs for the ApA junction, similar to the ribosomal system. The system with the highest yield is the UpA junction, with 20% yield at 4 ^{o}C at 7 d, and 42% at 25 ^{o}C at 7 d. In the cases where at least 10% ligation has occurred at 4 ^{o}C (i.e. Figure 32A ApG, Figure 32B UpA, and ApG), increasing the reaction temperature to 25 ^{o}C leads to a two-fold increase in ligation yields.

The difference in ligation yields between the ribosomal and the Watson-Crick systems could be as a result of the presence of mismatch at the ligation junction for the 5´-UAA-3´/5´-GAA-3´ system. The difference within each system, however, highlights the sequence dependence of cyclic phosphate ligation, the advantage of base-pairing, and possible nearest neighbor effects. In Figure 32A, for the lowest yielding ligation junction ApA, it is observed that the A's are both mismatch pairs. This mismatch could contribute to instability at the ligation nick, preventing the cyclic phosphate from attaining favorable conformation for ligation. It is possible that the position of the mismatch can play a key role in determining the efficacy of ligation. For example, in Figure 32A, for the UpA junction, the mismatch is at the nucleophile and at its nearest neighbor, and only 5% ligation yield is observed. On the contrary, the ApG junction in Figure 32A, has a mismatch at the cyclic phosphate, and its nearest neighbor, and the yield increases to 18%.

Note that the absence of a mismatch does not lead to uniformity in the ligation results, as shown in Figure 32B. In this case, all of the junctions are Watson-Crick base-pairs, still a difference in ligation yields ranging from 20% to 40% is observed. Several factors could be contributing to the difference in the ligation yields that are observed, including the reactivity of the cyclic phosphate and nucleotide junction (i.e. ligating between a U-A, and A-A, and an A-G), the stacking of the purines at the ligation nick and at the nearest neighbors, or the difference in the hydrolysis rate between the Up and Ap ligating junctions.

5.4 Kinetics of Ligation

To verify that 2 d and 7 d time points are good indicators of the instantaneous rates and maximum yields, longer term kinetic experiments were conducted, as shown Figure 33. In particular, a comparison between a 5´-UpAA-3´ and 5´-UAAp-3´ was made for both the ribosomal and Watson-Crick templates.

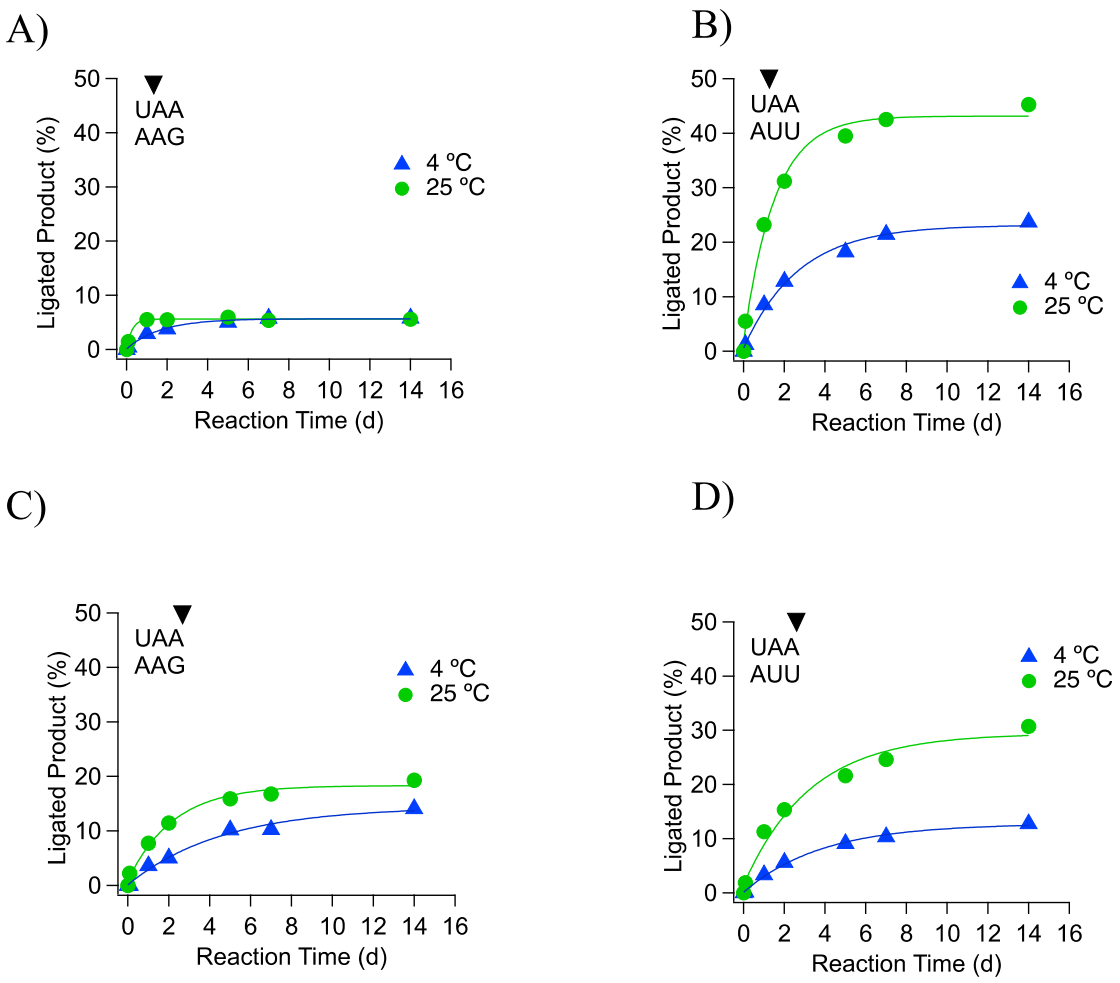

Figure 33. Kinetics of ligation of the 5´-UpAA-3´ and 5´-UAAp-3´ for the GAA and UUA ligation template systems. (A) Kinetics of 5´-UpAA-3´/5´-GAA-3´ (B) Kinetics of 5´-UpAA-3´/5´-UUA-3´ (C) Kinetics of 5´-UAAp-3´/5´-GAA-3 (D) Kinetics of 5´-UAAp-3´/5´-UUA-3´. Data points were obtained as described in Figure 31 caption. Markers represent the experimental data while the solid line is an exponential fit of the data. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. The FAM substrate and the template were in equimolar amounts of 1 µM. The black triangles represent the cyclic phosphate position.

In Figure 33A, for the 5´-UpAA-3´/5´-GAA-3´ system a maximum ligation yield of 5% is observed at both 4 °C and 25 °C. At 4 °C, this maximum is reached after 5 d, whereas at 25 °C, the maximum is attained after 2 d. Further incubation of the substrates does not lead to an increase in yield of the products. On the contrary, for the 5´-UpAA-

3′/5′-UUA-3′ system in Figure 33B, there is a difference in ligation rates and yields between the reaction temperatures. A nearly two-fold difference between the yields at 4 °C and 25 °C is observed. For example, at 2 d, the ligation yield increases from 15% to 31% as temperature increases from 4 °C and 25 °C, while at 14 d the yield increases from 21% to 42% at 4 °C and 25 °C respectively. After 7 d the yield appears to have approached its maximum and increases only slightly at 14 d.

In Figures 33C and D, ligation yield is compared for the 5′-UAAp-3′/5′-GAA-3′ and 5′-UAAp-3′/5′-UUA-3′ system respectively. In Figure 33C, 12% yield is obtained after 14 d at 4 °C, and this increases to 20% yield when the temperature is increased to 25 °C. This yield reflects a two-fold and four-fold increase compared to what was observed in Figure 33A at 4 °C and 25 °C respectively. Surprisingly, in Figure 33D, the same rates and yields are observed for the UAAp-3′/5′-UUA-3′ system at 4 °C, with both the ribosomal and Watson-Crick systems attaining 12% yield after 14 d at 4 °C. At 25 °C, a two-fold increase in ligation yields is observed similar to Figure 33B, with maximum ligation of 30% occurring after 14 d. It is worth noting that the reactions in Figures 33C and D appear to have not yet reached a plateau, unlike their counterparts in Figures 33A and B.

The difference in ligation rates and yields between the 5′-UpAA-3′ and 5′-UAAp-3′ could be as a result of the different rates of hydrolysis, as stated in the previous section. This hydrolysis will lead to a depletion of the amount of cyclic phosphate substrate available for ligation, especially after long incubation times such as 14 d, and will be further investigated in the next section. A second reason for the difference in rates and yields between the two systems could be as a result of thermal stability. In order to rule this possibility out, thermal denaturation studies were conducted, and a sample measurement is

shown in Figure 34. The results of this studies, shown in Table 1, revealed that the phosphate substrate strand had a melting temperature of either 65 ºC, for the 5´-UpAA-3´ or 70 ºC for the 5´-UAAp-3´ system, with no difference when either the 5´-GAA-3´ or 5´-UUA-3´ template was used. Therefore, the difference in reactivity in Figures 32 and 33 cannot be attributed to the T_m's.

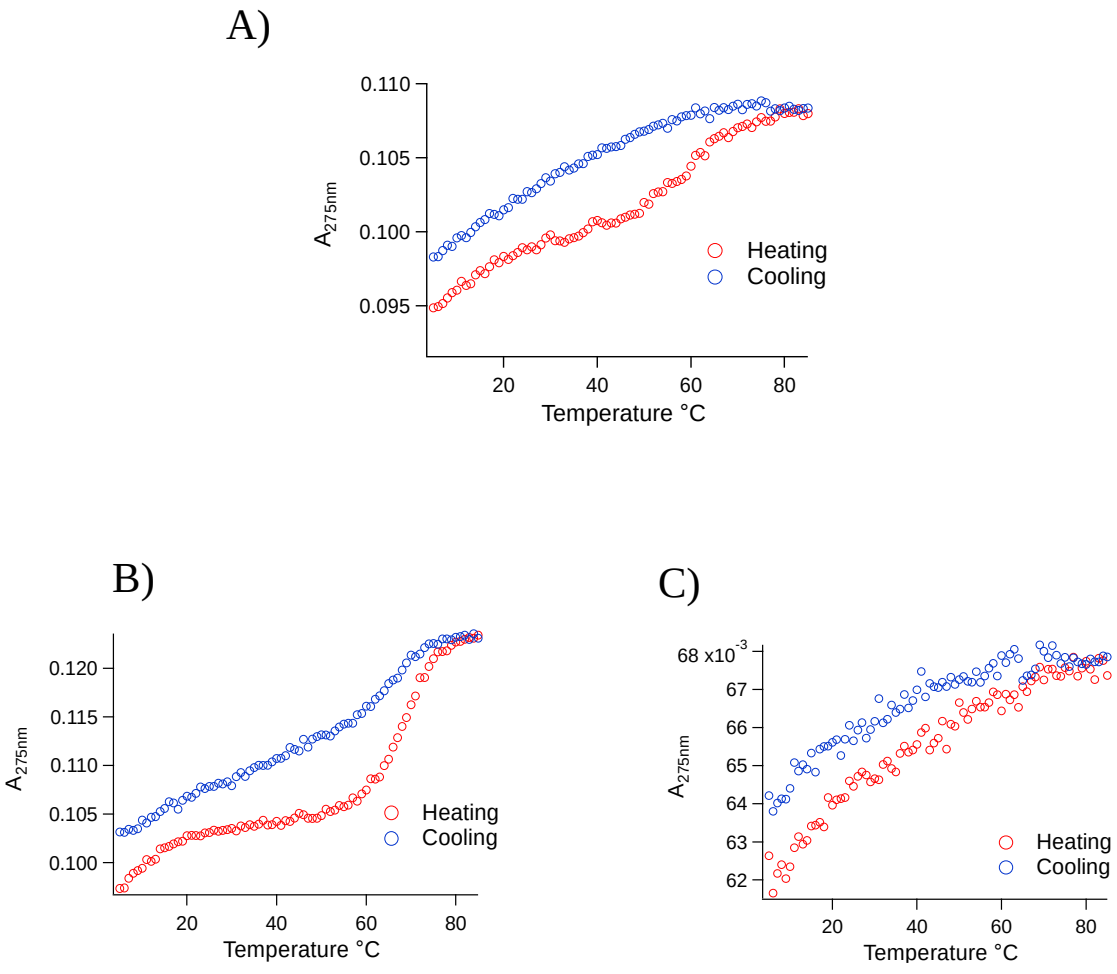

Figure 34. Determination of melting temperature (T_m) of the template-substrate assembly for the 5´-UpAA-3´/5´-UUA-3´ system by monitoring absorption at 275 nm as a function of temperature. (A) The 5´-UUA-3´ template shows a small change in absorption at 50 ºC. (B) The Up (i.e. the 3´-phosphate) substrate and the 5´-UUA-3´ template are monitored. The substrate has a T_m of 65 ºC. (C) The FAM substrate and the 5´-UUA-3´ template are monitored. No clear transition of the substrate can be seen.

Table 6. Summary of melting temperature measurements for the 5´-UpAA-3´ and 5´-UAAp-3´ for both the GAA and UUA ligation template systems. Measurements were conducted in buffer containing 100 mM NaCl, and 2.5 mM Tris, pH 8.3, and 10 mM $MgCl_2$.

	Melting Temperatures (T_m) ^oC		
	Template	3´-p substrate + template	FAM substrate + template
5´-UpAA-3´/5´-GAA-3´	55	65	< 20
5´-UpAA-3´/5´-UUA-3´	N/A	65	< 20
5´-UAAp-3´/5´-GAA-3	55	70	< 20
5´-UAAp-3´/5´-UUA-3	N/A	70	< 20

Data in Table 6 also showed a high T_m of 55 ^oC for the 5´-GAA-3´ template, which could have an impact on substrate's ease of accessibility to the ligation nick. Therefore, new templates were designed to remove this structure and a subsequent decrease in T_m from 55 ^oC to 20 ^oC was observed for the template. However, subsequent ligation studies revealed that the decrease in T_m had no impact on ligation yields, shown in Figure 35.

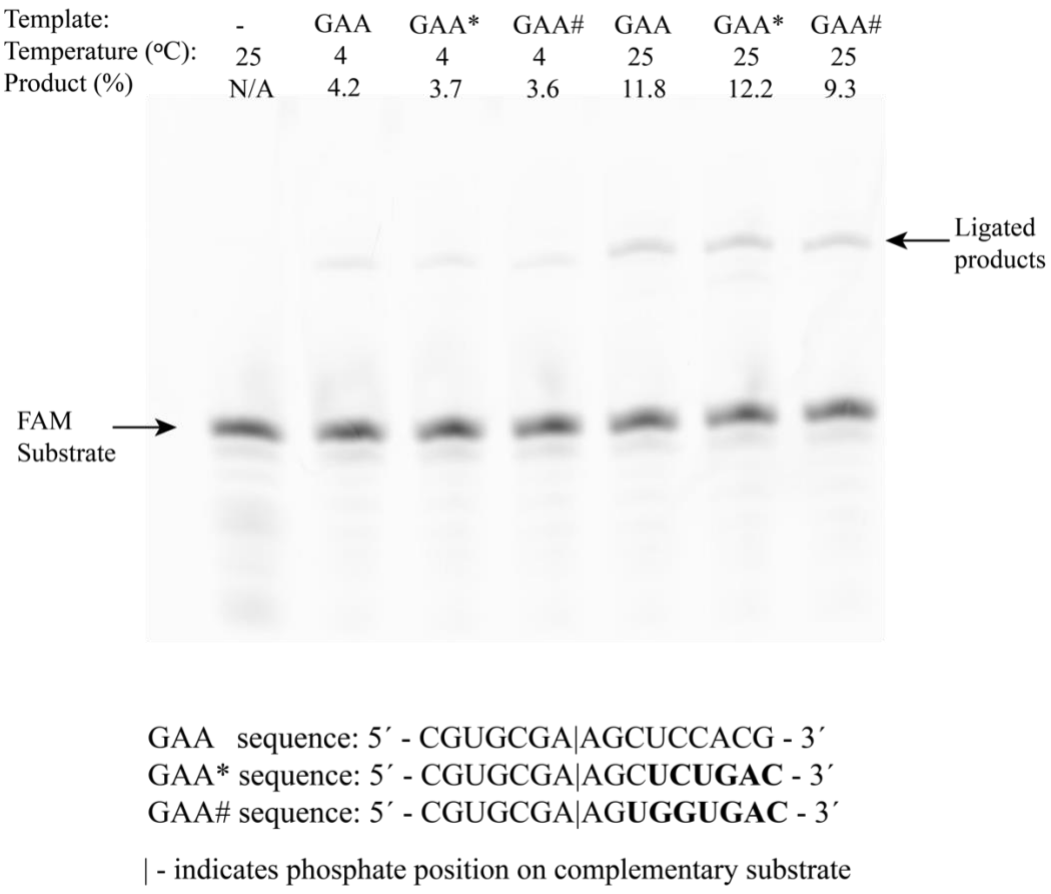

Figure 35. Ligation yields of ribosomal GAA templates with varying sequence upstream of the cyclic phosphate. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. Reaction was quenched after 7 d.

5.5 The perils of cyclic phosphate hydrolysis

Given the long reaction times studied in Figure 32 and Figure 33, it is expected that the rate of hydrolysis might also contribute to the limit on the reaction yield, especially at the higher reaction temperature of 25 °C. To this effect, the cyclic phosphate substrate from the 5′-UpAA-3′ and 5′-UAAp-3′ systems were incubated in the reaction buffer, in the absence of template and the FAM substrate, as shown in Figure 36A. After incubation, the

template and the FAM substrate were added to the reaction mix and allowed to react for up to 2 d.

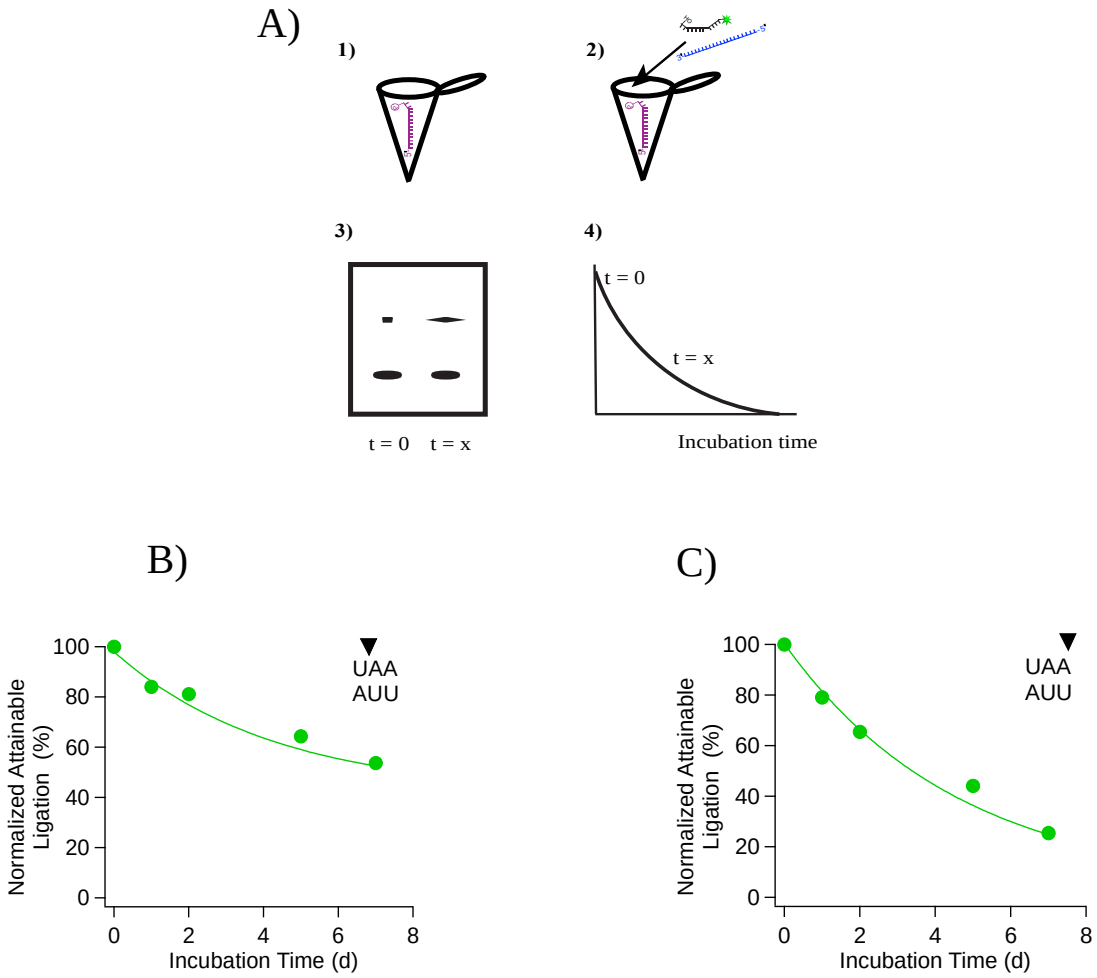

Figure 36. Hydrolysis of the cyclic phosphate substrate over time. (A) Schematic of the hydrolysis reaction. 1. Incubation of 2 µM the cyclic phosphate substrate in reaction buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl for the incubation time period. (2) Addition of 1 µM of the FAM substrate and the template, after which the ligation reaction proceeds for 2 d. (3) Gel electrophoresis of the different timepoints. (4) Plot showing comparison of the reaction product at time t = x to products formed at time t = 0. B) Ligation of 5´-UpAA-3´/5´-UUA-3´ C) Ligation of 5´-UAAp-3´/5´-UUA-3´. All reactions were carried out 25 °C. Data points were obtained as described in Figure 31 caption. Markers represent the experimental data while the solid line is an exponential fit of the data to guide the eye. The black triangles represent the cyclic phosphate position.

Figure 36B and 36C show the results for the 5´-UpAA-3´/5´-UUA-3´ and the 5´-UAAp-3´/5´-UUA-3´system respectively. The normalized yield after 2 d of reaction for a time zero incubation is shown as a 100%. As the cyclic phosphate in incubated for up to 7 d a decrease in the amount available for ligation is expected due to cyclic phosphate hydrolysis (Figure 30), as observed in Figures 36B and 36C.

In Figure 36B, after 2 d incubation, 80% of attainable ligation is observed at 25 ^{o}C. As incubation increases to 7 d, the attainable ligation yield falls further to 52 % yield. Since all the reactions were ligated for 2 d after incubation, the dramatic decrease in yield is due to cyclic phosphate hydrolysis i.e. 20 % of substrate is lost to hydrolysis after 2 d and 48 % after 7 d. As a result, it is no surprise that the maximum ligation that can be obtained under our regular reaction conditions is ~50 % yield, as shown in Figure 33B.

Figure 36C shows that after 2 d incubation, 60% attainable ligation is obtained at 25 ^{o}C for the 5´-UAAp-3´/5´-UUA-3´system. After incubation of up to 7 d, the attainable ligation yield is reduced to 20%. These results imply that over 80% of the cyclic phosphate substrate is hydrolyzed after 7 d, therefore, a yield of ~20% yield is expected after 7 d reaction as observed for the 25 ^{o}C data in Figure 33D.

The lower attainable yields observed for the 5´-UAAp-3´/5´-UUA-3´compared to the 5´-UpAA-3´/5´-UUA-3´ system demonstrates that hydrolysis of the former is occurring at a ~1.5x faster rate than the latter. This difference in hydrolysis rate explains why there is a large difference in ligation yields in Figure 33B and 33D despite the fact that they are both Watson-Crick paired system. Note that in Figure 32B, the yields of the substrates with a 5´-A- cyclic phosphate are similar despite the differences in the ligation junction. Lastly,

control experiments confirmed that in the absence of Mg^{2+} prior to ligation, a minimal amount of hydrolysis occurs, similar to finding by others (Table 7) (113).

Table 7. Raw data of hydrolysis ligation yields. The normalized yields are represented in Figure 36.

	5´-UpAA-3´/5´-UUA-3´		5´-UAAp-3´/5´-UUA-3´	
Incubation Time (d)	Ligation yield (%)	Normalized yield (%)	Ligation yield (%)	Normalized yield (%)
0	29.3	100	13.3	100
1	24.6	84	10.5	79
2	23.7	81	8.7	65
5	18.9	64	5.9	44
7	15.7	54	3.4	25
7 – no Mg^{2+}	24.5	84	10.1	76

5.6 Identity of cyclic phosphate base-pair and effects on ligation yields

Following up on the observations that mismatches could be incorporated easily at certain cyclic phosphates (Figure 32), the relationship between mismatch types and ligation yields was investigated. In Figure 37, two primary types of systems were investigated, one in which a 5´-A-OH nucleophile was used and another in which a 5´-C-OH nucleophile was used. The nucleophile is Watson-Crick paired, while the cyclic phosphate base will

either be a Watson-Crick bp or a mismatch bp. Given that the rate of hydrolysis among the four cyclic phosphate bases could vary widely, comparisons will be made solely between systems in which cyclic phosphate base is similar.

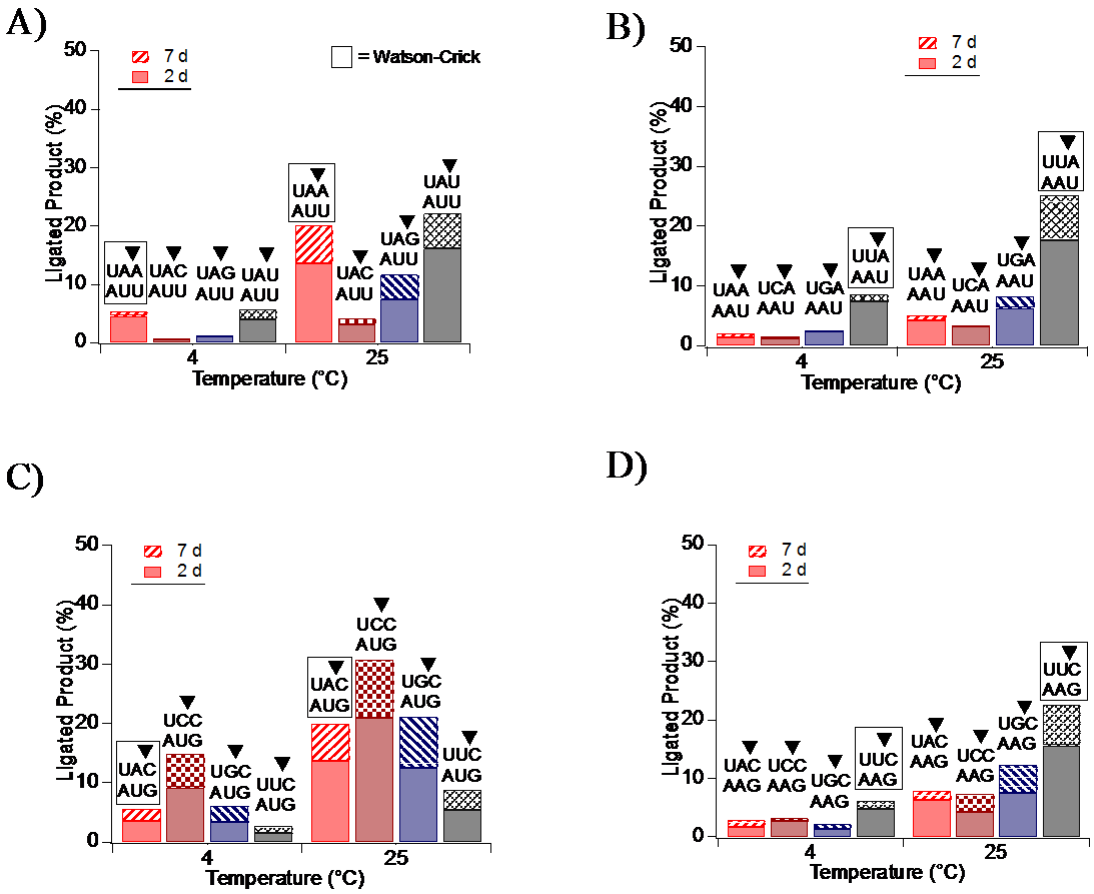

Figure 37. Ligation yields of cyclic phosphate substrates compared to 4 templates. (A) Ligation of cyclic phosphates on a 5´-UUA-3´ template. (B) Ligation of cyclic phosphates on a 5´-UAA-3´ template. (C) Ligation of cyclic phosphates on a 5´-GUA-3´ template. (D) Ligation of cyclic phosphates on a 5´-GAA-3´ template. Watson-Crick base-pairs are enclosed in the black rectangle. All reactions were carried out for 2 d and 7 d indicated by the unfilled and shaded bars respectively, at 4 °C, and 25 °C. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. The FAM substrate and the template were in equimolar amounts of 1 µM. The directionality of the labeled substrates is 5´-UAA-3´ (top label) /5´-UUA-3´ (bottom label). Data points were obtained as described in Figure 31 caption.

Figures 37A and 37B shows ligation yields when the cyclic phosphate base-pair (bp) is either a U or an A. In both cases the maximum ligation yield is ~22% for the U bp, or 25% for the A bp, and occurs for the Watson-Crick bp system. Figure 37A, reveals yields of ~5% to 22% for the mismatch bp, with the U-U mismatch bp showing the least ligation yield and the C and G incorporating as well as the A-U Watson-Crick bp. In contrast, Figure 37B demonstrates that the A-A, G-A, and C-A mismatch bp only attain ~8% yield for the best systems. These yields for the incorporation of the A mismatch trail far behind the comparative Watson-Crick system which has a yield of 25%. It is no surprise that the G-U and C-U mismatch have a ligation yield that is similar to the Watson-Crick A-U bp, given that G-U wobble base pairs and C-U mismatches are known to be stable within the duplex (114-116).

Figures 37C and 37D draw a parallel comparison to Figures 37A and 37B, except in this case a 5´-C-OH nucleophile is used. The maximum ligation yields in Figures 37C and 37D range from 30% to 20%. Similar to the earlier findings, the incorporation of a U mismatch is 10-15% higher than the incorporation of an A mismatch. Figure 37C shows the G-U mismatch and the A-U Watson-Crick bp attain the same ligation yield of 20% for the best ligation conditions. Surprisingly, the C-U mismatch surpasses its Watson-Crick counterpart by 10% with a maximum ligation yield of 30%. This large difference in yield of the C-U mismatch compared to the Watson-Crick could be as a result of favorable stacking of the cytosine rings and the stability of the C-U bp within the duplex (115-117). Like observed in Figure 37B, Figure 37D shows poor incorporation of mismatches with an A bp.

Altogether, the results in Figure 37 reveal the ease of incorporation of some mismatch bp over others, with a U bp being favored over an A bp as indicated by the higher yields obtained with a U mismatch in Figures 37A and 37C. Holistically, there does not seem to be a difference associated with the identity of the nucleophile, as the same ligation behavior was observed for both the 5´-C-OH and 5´-A-OH nucleophiles. Unlike Figure 32B, in which there was a contrast between ligation yields and rates for two of the systems, all the Watson-Crick bp in the systems in Figure 37 have a ligation yield of ~20%. The similarity between the Watson-Crick bp systems in Figure 37 could be due to the fact the ligation nick has the same flanking sequence, in contrast to Figure 32B, in which the flanking sequences at the ligation nick was different for each system.

Lastly, as a proof of principle showing how flexibility at the ligation nick can allow for ligation, a system in which the cyclic phosphate junction was following by an adjacent missing bp across the ligation nick was designed. These results, shown in Figure 38, demonstrate the ease with which an unpaired U pops out of the ligation site compared to an unpaired A, with a maximum ligation yield of 20% for the best ligating system.

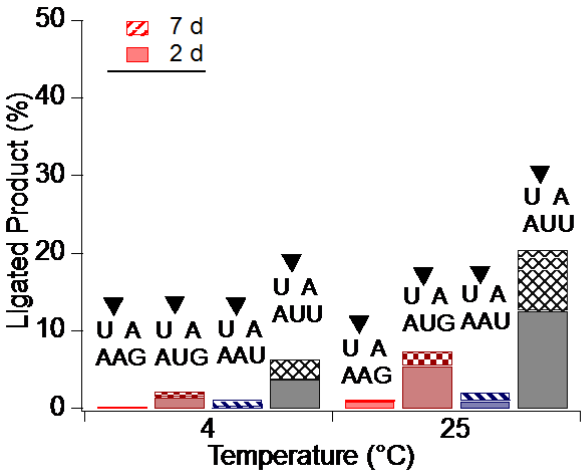

Figure 38. Ligation in loops. All reactions were carried out for 2 d and 7 d indicated by the unfilled and shaded bars respectively, at 4 °C, and 25 °C. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. The FAM substrate and the template were in equimolar amounts of 1 µM. The directionality of the labeled substrates is 5´-UAA-3´ (top label) /5´-UUA-3´ (bottom label).

5.7 Mismatch incorporation at the nucleophile base-pair

The results in the previous section compelled us to compare the ease of incorporation of a mismatch at the nucleophile to that at the cyclic phosphate. In this system design, a 5´-A- cyclic phosphate will be used to ligate four different nucleophiles. Two types of base-pair mismatches will be studied, a mismatch with a U bp, and a mismatch with an A bp.

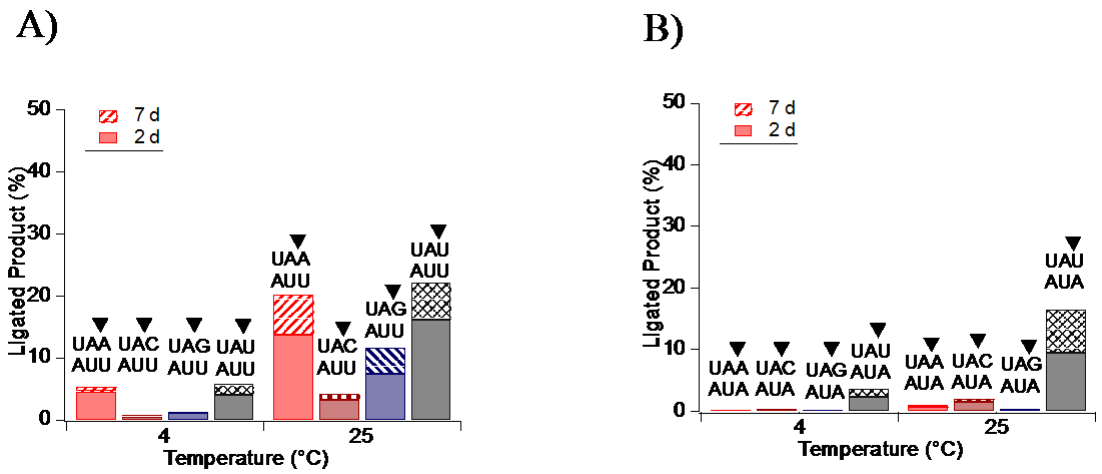

Figure 39. Ligation yields of different nucleophiles compared to 2 templates. (A) Ligation on a 5´-UUA-3´ template. (B) Ligation on a 5´-AUA-3´ template. All reactions were carried out for 2 d and 7 d indicated by the unfilled and shaded bars respectively, at 4 °C, and 25 °C. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. The FAM substrate and the template were in equimolar amounts of 1 µM. The directionality of the labeled substrates is 5´-UAA-3´ (top label) /5´-UUA-3´ (bottom label).

In Figure 39A, the maximum ligation yields for the best performing system is ~22% yields, for both the Watson-Crick A-U bp and a mismatch U-U bp. The nucleophile mismatch bp C-U and G-U have maximum yields of 4% and 10% respectively. This is in contrast to what was observed for the cyclic phosphate mismatch C-U and G-U in Figure 37A, where the yields were similar to the Watson-Crick bp. The low yields observed in Figure 39A for the C-U and G-U mismatch supports our earlier observation that a mismatch at the nucleophile can be detrimental to yields, as seen in the ribosomal system in Figure 32A.

The U nucleophile was different from the other mismatch nucleophiles studied. Given the low yields observed with a U-U cyclic phosphate mismatch, it was expected that there will be little to no yields for the nucleophile mismatch. Surprisingly, Figure 39A

showed that its maximum ligation was similar to the Watson-Crick counterpart. Further experiments conducted with the U-U nucleophile mismatch, shown in Figure 40A and 40B confirm that it incorporates better than a U-U cyclic phosphate mismatch.

Figure 39B shows mismatch pair patterns that are identical to what was observed in Figure 37, with the U bp nucleophile showing general higher ligation yields than the A bp mismatch (comparing Figure 39A and Figure 39B). The Watson-Crick bp outcompetes the other ligation systems with a yield of 16% while less than 5% yields are observed for the mismatch base-pairs. Altogether, the low yields observed in Figures 39 and 40 show that mismatches at the nucleophile are tolerated to a lesser degree than when compared to cyclic phosphate mismatches.

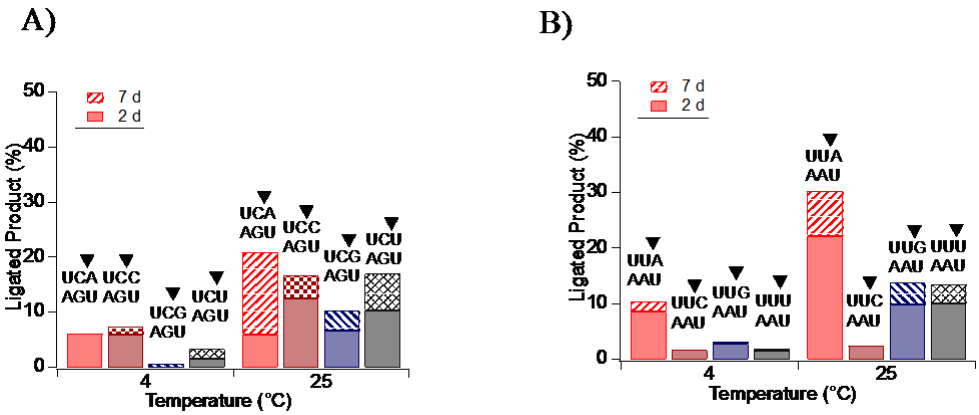

Figure 40. Ligation yields of the C and U cyclic phosphates for different nucleophiles. (A) Ligation on a 5´-UGA-3´ template. (B) Ligation on a 5´-UAA-3´ template. All reactions were carried out for 2 d and 7 d indicated by the unfilled and shaded bars respectively, at 4 °C, and 25 °C. The cyclic phosphate substrate: FAM substrate ratio was 2:1 in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. The FAM substrate and the template were in equimolar amounts of 1 µM.

5.8 Determining linkage type for newly formed products

One last question remains in the analysis of the cyclic phosphate ligation, the question of linkage type. Studies have shown the prevalence for the formation of non-canonical 2′-5′ linkages (27,106), over the canonical 3′-5′ linkages in cyclic phosphate ligation. Therefore, it was imperative that we investigate linkage type in our study. The best performing Watson-Crick systems, shown in Figure 32 were used as the sample set for these investigations, given their high ligation yields.

Nuclease P1 was used to digest the reactions, given its known ability to preferentially digest 3′-5′ linkages over 2′-5′ linkages and leave behind 5′-phosphate mononucleotides and dinucleotides (118-120). Control reactions were run with a template, product control such that only 3′-5′ linkages were present, and FAM substrates, as shown in Figure 41A and B. A clear separation of the four nucleotides and nucleosides can be seen.

In Figure 41C and 41D, the FAM substrates were enzymatically digested and run on the HPLC as controls, since most of our ligation products show over 80% of unreacted substrates. Mononucleotides can be seen, similarly to the template control in Figure 41A. There is also a large amount of Adenosine and Guanosine which can be seen at 7.8 mins and 10.2 mins respectively (Figure 41D). Lastly, as a proxy for the expected 2′-5′ linked dimers, a control 3′-5′ linked A-A dinucleotide was seen to elute at 12.3 mins (Figure 41 D).

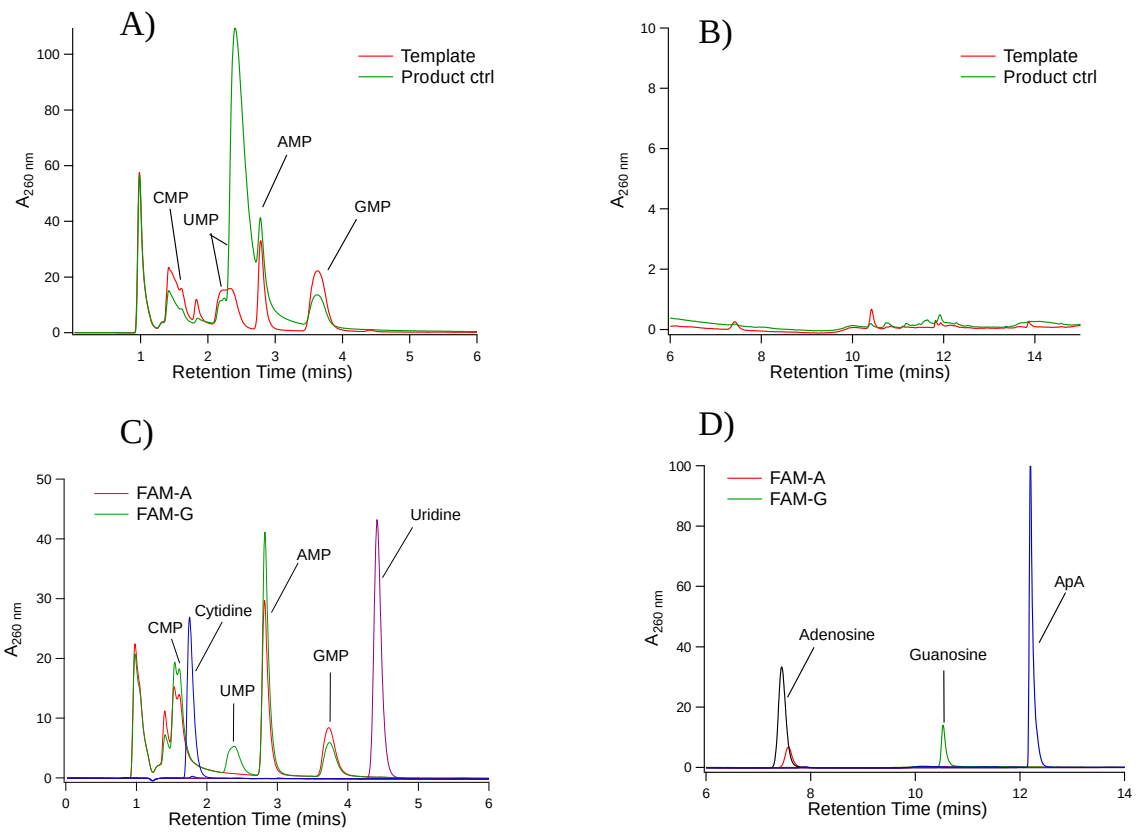

Figure 41. C18 HPLC traces of Nuclease P1 digested templates and FAM substrates used in the reactions in Figure 32. 5´-NMP and nucleoside standards were run alongside the digested products to verify their identity. (A) Digested 5´-UUA-3´ template run alongside an enzymatically ligated product known to product 3´-5´ linkages only. (B) Same plot as A shown for 6-14 mins retention time. (C) Digested FAM substrates for the 5´-UpAA-3´ and 5´-UAAp-3´ ligation system as shown in Figure 29. (D) Same plot as D shown for 6-14 mins retention time.

Given the successful separation of nucleotides and nucleosides after enzymatic digestion, the reaction products of the Watson-Crick system showed in Figure 32B were digested. The results of these cyclic phosphate ligated products are shown in Figure 42. A new peak, absent in the controls and presumed to be the 2´-5´ linked dimer, appears between 10-11 mins for the products investigated, and is labeled with an asterisk in Figure 42B. Thus confirming the preference for the formation of 2´-5´ linkages in the cyclic phosphate ligation over the canonical 3´-5´ linkages.

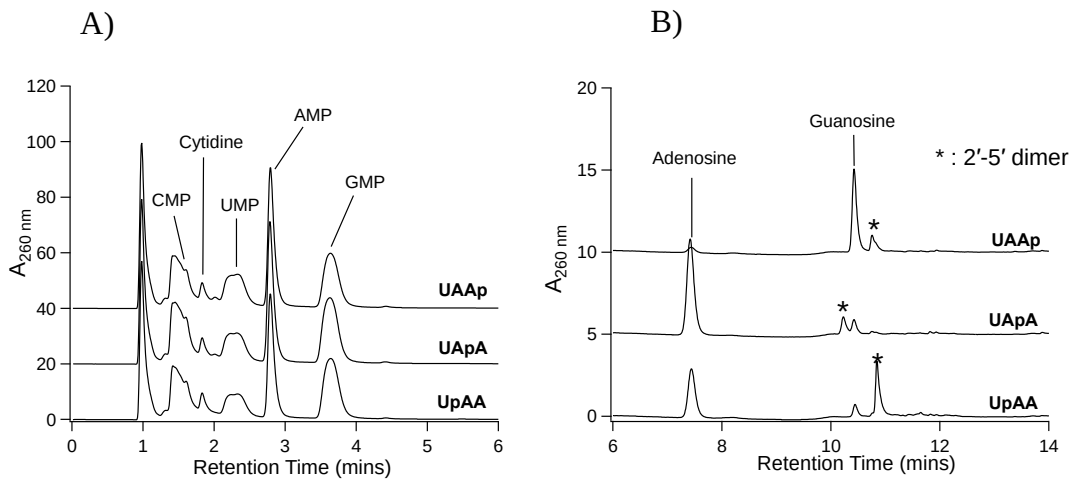

Figure 42. C18 HPLC traces of Nuclease P1 digested ligated products of Watson-Crick system in Figure 32B. The chemical ligation reaction was conducted at 25 °C. (A) Digested 5´-UAAp-3´, 5´-UApA-3´, and 5´-UpAA-3´ on a 5´-UUA-3´ template. (B) Same plot as A shown for 6-14 mins retention time.

5.9 Conclusion

Using the ribosomal 5´-UAA-3´/5´-GAA-3´ loop as a reference for ligation in loops, this Chapter set out to characterize the role of sequences in cyclic phosphate ligating systems. It is well documented that there could be enhanced cleavage in RNA loops (121,122), however, these are often site-specific and sequence dependent. Therefore, different potential sites for cleavage were chosen within the ribosomal loop that could lead to the formation of a cyclic phosphate. The results obtained in this study indicated that the most likely place for recombination within the ribosomal loop occurs when the cyclic phosphate was located after the A nucleotide at the 3´ end (i.e. a 5´-UAAp-3´/5´-GAA-3´ system). In contrast, when compared to the Watson-Crick base-pair system, we found that

the UpA junction (i.e. within the 5´-UpAA-3´/5´-UUA-3´ system) which has been shown to have higher cleavage rates than other nucleotides (123,124) gave the best ligation yields.

Given earlier work in Chapters 2-4 of this Thesis (125), it was expected that multiple factors would influence yields even in a controlled system. Therefore, thermodynamics and kinetics were investigated to see the influence on yields. Results in this study demonstrated that the substrates had similar stabilities with the *Tm*'s of the cyclic phosphate substrate indicating that it would be completely bound to the template at the reaction temperatures studied. Ligation rates were generally slow, often requiring up to 7 d for maximum ligation yields to be obtained. However, these low yields and slow reaction rates are similar to what was observed by others (82,83).

The systematic characterization of sequences in this Chapter revealed the extent to which certain types of mismatch base-pairs could be incorporated into the system. In general, high ligation yields were observed when a mismatch was present at the cyclic phosphate, compared to a mismatch at the nucleophile. It could be that the presence of a mismatch at the nucleophile reduces the ability to attain the in-line conformation required for ligation (24,51,105), unlike a cyclic phosphate mismatch. Lastly, cyclic phosphate hydrolysis was found to be a significant source of substrate loss in these studies, which would render this ligation method unsustainable in the prebiotic world.

5.10 Materials and Methods

5.10.1 Oligonucleotide preparation and verification of cyclic phosphate formation

All oligonucleotides, except for the fluorescein labeled oligonucleotides, were purchased from Integrated DNA Technologies (IDT) where they were synthesized using standard phosphoramidite methods. Upon receipt from IDT, the oligonucleotides were resuspended in 18.2 MΩ/cm water (Barnstead NanopureTM). The fluorescein labeled oligonucleotides were purchased from Horizons Discovery and were similarly resuspended as the IDT oligonucleotides.

The 3´p substrates were quantitatively converted to cyclic phosphates by incubating with 50 mM (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide) EDC (Sigma-Aldrich) and 100 mM MES pH 6.0 at 37 ºC for 40 mins. After the formation of cyclic phosphates ethanol precipitation was performed and the oligonucleotides were resuspended in 18.2 MΩ/cm water.

The cyclic phosphate intermediate was verified by HPLC traces shown in Figure 43.

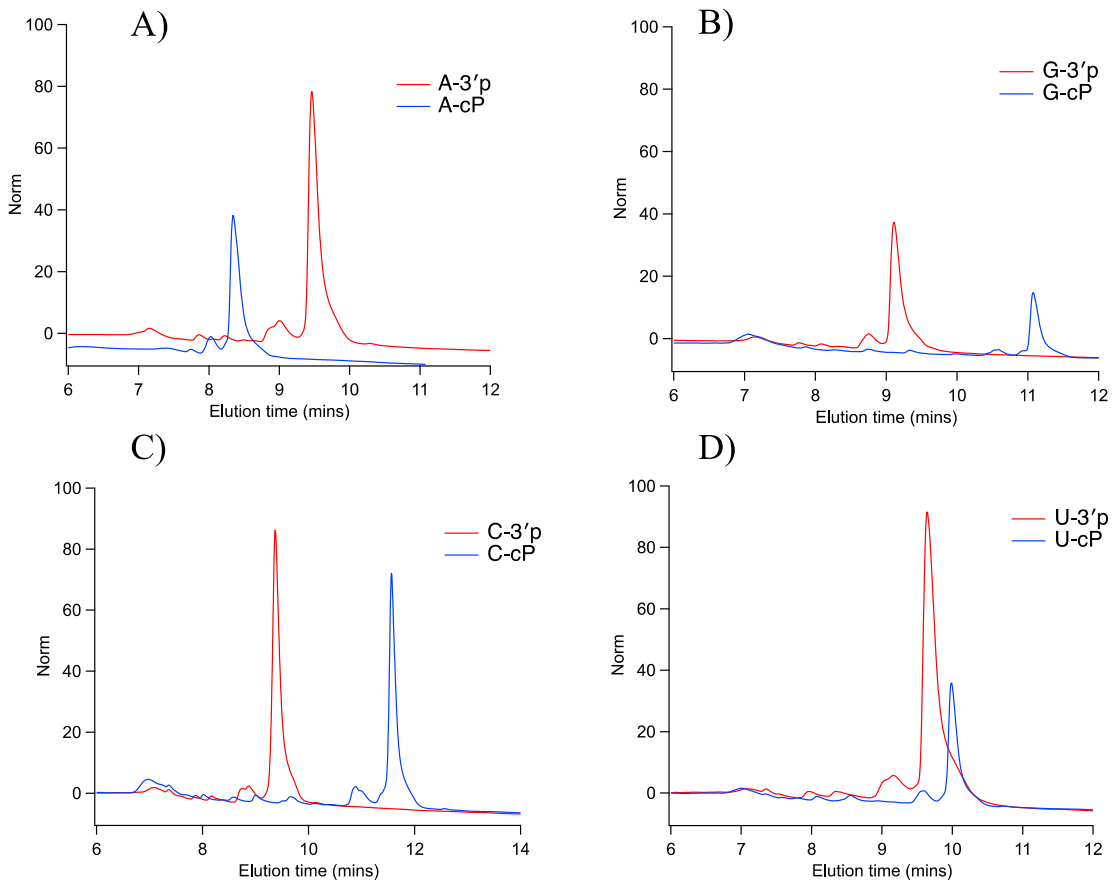

Figure 43. HPLC traces of cyclic phosphate substrates used in Figure 37. The details of the HPLC method can be found in *Section 4.8.3*.

5.10.2 Oligonucleotide sequences

Sequences	Position in Thesis
Cyclic P substrate: 5´-CGUGGAGCUUp-3´ **FAM substrate**: 5´-AGCACGACAC-3´-FAM **Template**: 5´-CGUGCUAAGCUCCACG-3´	Figure 31
Cyclic P substrate: 5´-CGUGGAGCUp-3´	

FAM substrate: 5´-AAGCACGACAC-3´-FAM **Cyclic P substrate**: 5´-CGUGGAGCUAp-3´ **FAM substrate**: 5´-AGCACGACAC-3´-FAM **Template**: 5´-CGUGCGAAGCUCCACG-3´ **Template**: 5´-CGUGCUUAGCUCCACG-3´ **Cyclic P substrate**: 5´-CGUGGAGCUAAp-3´ **FAM substrate**: 5´-GCACUGACAC-3´-FAM **Template**: 5´-CAGUGCGAAGCUCCACG-3´ **Template**: 5´-CAGUGCUUAGCUCCACG-3´	Figure 32, Figure 33, Figure 36, Table 6, Table 7
Cyclic P substrate: 5´-CGUGGAGCUAp-3´ **Cyclic P substrate**: 5´-CGUGGAGCUCp-3´ **Cyclic P substrate**: 5´-CGUGGAGCUGp-3´ **Cyclic P substrate**: 5´-CGUGGAGCUUp-3´ **FAM substrate**: 5´-AGCACGACAC-3´-FAM **Template**: 5´-CGUGCUUAGCUCCACG-3´ **Template**: 5´-CGUGCUAAGCUCCACG-3´ **FAM substrate**: 5´-CGCACGACAC-3´-FAM	Figure 37

Template: 5´-CGUGCGUAGCUCCACG-3´ **Template**: 5´-CGUGCGAAGCUCCACG-3´	
Cyclic P substrate: 5´-CGUGGAGCUp-3´ **FAM substrate**: 5´-AGCACGACAC-3´-FAM **Template**: 5´-CGUGCGUAGCUCCACG-3´ **Template**: 5´-CGUGCGAAGCUCCACG-3´	Figure 38
Cyclic P substrate: 5´-CGUGGAGCUAp-3´ **Cyclic P substrate**: 5´-CGUGGAGCUCp-3´ **Cyclic P substrate**: 5´-CGUGGAGCUUp-3´ **FAM substrate**: 5´-AGCACGACAC-3´-FAM **FAM substrate**: 5´-CGCACGACAC-3´-FAM **FAM substrate**: 5´-GGCACGACAC-3´-FAM **FAM substrate**: 5´-UGCACGACAC-3´-FAM **Template**: 5´-CGUGCUUAGCUCCACG-3´ **Template**: 5´-CGUGCAUAGCUCCACG-3´ **Template**: 5´-CGUGCUGAGCUCCACG-3´ **Template**: 5´-CGUGCUAAGCUCCACG-3´	Figure 39, Figure 40

5.10.3 Ligation experiments

For a standard reaction mix, the template and the fluorescently labeled (FAM) substrate were present at 1 μM along with the cyclic phosphate substrate oligonucleotides which was present at 2 μM, in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. This reaction mix was then incubated at the specified reaction temperature. After the indicated reaction time, 5 μL of each reaction was quenched by the addition of 1 μL of 100 mM EDTA and 6 μL of 2x loading buffer and dye (95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, 5 mM EDTA pH 8.0). Each sample was then stored at -80 °C until analyzed by denaturing polyacrylamide gel electrophoresis.

Polyacrylamide gels (Fisher BioReagents™ acrylamide/bis-Acrylamide 29:1, 40% solution) were 20% denaturing gels (8 M urea) run in 1x TBE (Tris, Boric Acid, and EDTA pH 8.0) buffer, 16 cm wide X 16 cm long. Gels were pre-run at 14 W and 300-400 V for at least 30 min prior to loading. Samples were run at the same conditions for 1 h. Imaging was done using a Typhoon Trio+ laser scanner (GE Healthcare) at a resolution of 50 μm and with a photomultiplier setting between 300-500. 'FAM filter' images were acquired using the 'FAM channel', which refers to 488 nm excitation and a 526 nm emission filter. Densitometry analysis was performed using utilities within the ImageJ software package (NIH), as discussed in *section 2.9.5*.

5.10.4 Melting temperature determination

UV absorbance was used to monitor the thermal denaturation of the template and substrates. RNA samples were prepared with varying concentrations in buffer containing 100 mM NaCl, and 2.5 mM Tris, pH 8.3, and 10 mM $MgCl_2$. UV measurements were performed on 1 mm quartz cuvettes in a temperature-controlled UV-Vis spectrophotometer (Agilent 8453) with nitrogen flowing through the sample chamber at low temperatures. To determine T_m values, heating and cooling traces were generated for each sample by recording spectra (220-400 nm) from 5 to 85 ^oC at intervals of 1 ^oC. Melting curves were generated using the signal at a single wavelength. For UV- Vis studies, the absorbance change at 260 nm or 275 nm was used. T_m values were determined as described by Mergny & Lacroix (68).

5.10.5 Hydrolysis experiments

2 μM cyclic phosphate substrate were incubated for a given length of time at the specified reaction temperature in a buffer containing 10 mM $MgCl_2$, 2.5 mM Tris, pH 8.3 and 25 mM NaCl. Afterwards, 1 μM of the FAM substrates and the templates were added to the reaction mix and the ligation reaction proceeded for 2 days. After 2 days, 5 μL of each reaction was quenched by the addition of 1 μL of 100 mM EDTA and 6 μL of 2x loading buffer and dye (95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, 5 mM EDTA pH 8.0). Each sample was then stored at -80 ^oC until analyzed by denaturing polyacrylamide gel electrophoresis.

The amount of product obtained for a sample in which no prior incubation occurred before the ligation reaction (time zero) was normalized to a 100% attainable yield. All other product yields were compared to the yield at time zero.

5.10.6 Digestion experiments

Ligation product standards were obtained enzymatically. Afterwards, the chemical ligation and enzymatic ligation product were digested using nuclease P1 purchased from New England Biolabs (NEB). The reaction was incubated at 37 ^{o}C for 40 mins, after which it was quenched by the addition of 2 μL of 0.5 M EDTA. The samples were run on a C-18 HPLC column described below, and a sample gel showing after nuclease digestion is sown in Figure 44.

HPLC analyses were conducted on an Agilent 1260 Infinity HPLC. Reaction products were separated using a Kinetex XB-C18 column (150 × 2.1 mm, 2.6 μm particle size). The flow rate was 0.3 mL/min and the column temperature was held at 25°C.

Eluent A is made of 0.1% formic acid; Eluent B is made of 100% Acetonitrile. 0-5 minutes isocratic flow at 100% (A). 5-25 mins used gradients from 0 – 55% (B). 26-36 mins 55-100 % (B). 37-50 mins 100% (A). Elution was recorded at 220 and 260 nm wavelengths, with a 180–400 nm spectrum detected in 2 nm steps.

To characterize reaction products, NMP standards were spiked into product mixtures.

Select samples were then further characterized by liquid chromatography mass spectrometry using an Agilent 1290 HPLC; Agilent 1260 Autosampler and DAD UV–Vis detector; path length: 0.6 cm; Agilent 1260 quaternary pump and RID; column: Phenomenex Kinetex 2.6 mmxB-C18100Å, LC column 150 × 2.1 mm; column temp: 25°C; 10 μl injection with needle wash, 100 μl s^{-1} injection speed. The buffer and eluent speeds were identical to that used on the Agilent 1260 Infinity HPLC, described earlier.

This system was coupled to an Agilent 6130 single quad MS Electrospray Ionization Mass Spectrometry system with scanning of ±65 to ±2,000 *m*/*z* and capillary voltage of 2.0 kV.

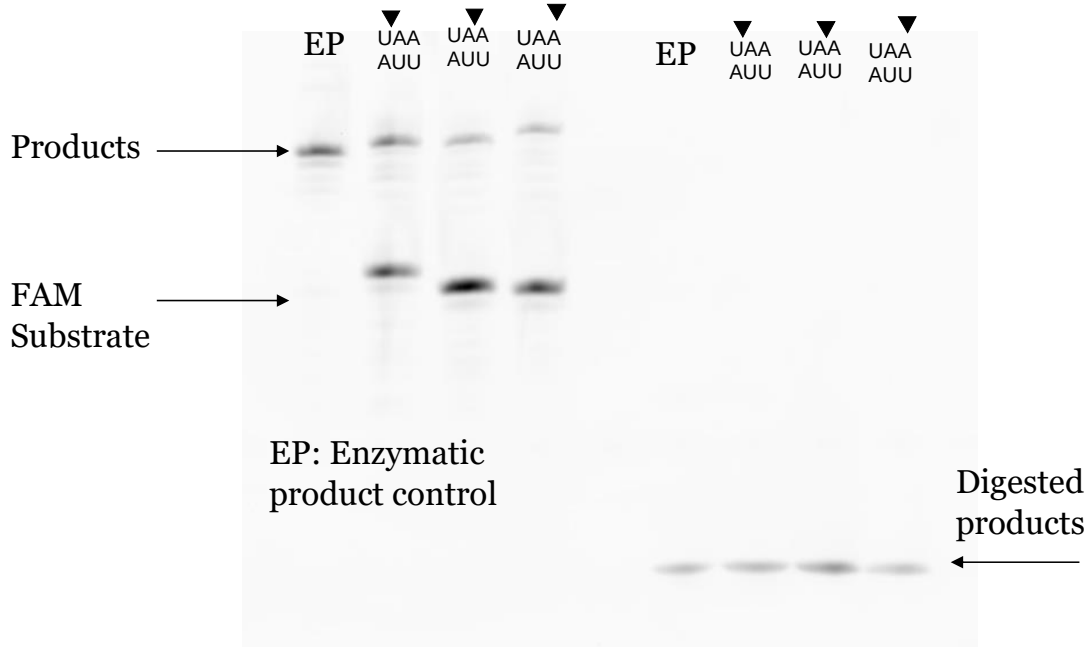

Figure 44. Sample gel electrophoresis image of Nuclease P1 reaction product digestion.

5.10.7 Optimization of buffer components

Prior to selecting reaction conditions for cyclic phosphate ligation, experimental optimization was performed for pH and magnesium concentrations. The results of these experiments are shown in Figure 45. At high reaction temperatures, the number of ligated products was similar for both pH 8.5 and pH 9.5. Similarly, there was no observable difference in yields when Mg^{2+} concentrations were increased.

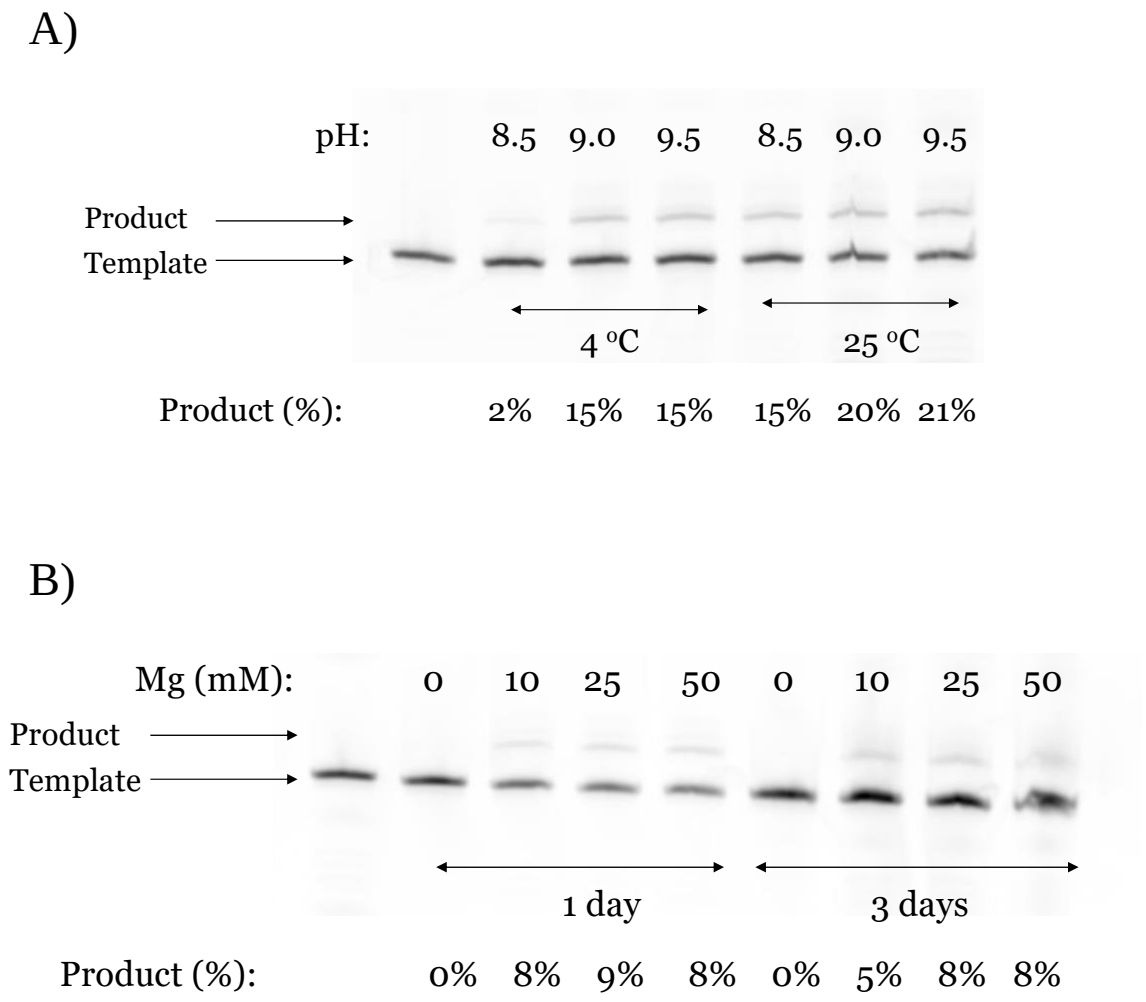

Figure 45. Optimizing reaction conditions for RNA ligation.

CHAPTER 6. CONCLUSIONS AND FUTURE OUTLOOK

6.1 Conclusions

Enzyme free replication has been the goal of prebiotic chemists for many decades (8,10), however, many of the proposed pathways to replication have fallen short of achieving a true replication system (126-129). One of the major challenges to developing a true replication system is the absence of a prebiotic route for ligation of short polymers that could have arisen from synthesis of monomers or via breakage of longer strands. If RNA (i.e. a true RNA molecule) was the first prebiotic polymer (5-8), then there would have existed a pathway for it to repair cleaved strands and evolve into complex structures.

Several methods have been proposed for enzyme free ligation, including the use of activated imidazole species (13-15), the use of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) (16,17,49), or by a combination of both methods (18-21). However, none of these methods have been widely accepted as a true solution to the problem given the difficult prebiotic synthesis of imidazole, and the implausibility of EDC as a prebiotic agent. The idea that RNA can recombine using a cyclic phosphate intermediate, is one that has gained traction in the last few years (22-25), but has not been developed further due to the low yields of product formation. Therefore, this dissertation set out to characterize the efficiency of non-enzymatic ligation and understand the limitations of the proposed ligation methods in a prebiotic context.

Work in this Thesis began by using DNA as a model system for ligation, given its similarities to RNA and with the expectation that knowledge derived from this study could

be transferred to the RNA system. In addition, since DNA did not form the intermediate cyclic phosphate, it was expected that it could represent a case for the most efficient RNA system. EDC was used as the activating agent in this Thesis because of its widespread use to generate a cyclic phosphate for RNA (22-25), and in the development of DNA nanotechnology (44).

One of the main findings of our DNA ligation studies was the need to select optimum reaction temperatures for a particular system. For example, high reaction temperatures were needed to increase ligation kinetics while low reaction temperatures ensured substrates remained bound to the templates prior to ligation. Conversely, high reaction temperatures were needed to allow optimum conformation of the activated phosphate prior to ligation, while low reaction temperatures were needed to prevent loss of substrates due to EDC modifications. Thus, creating a juxtaposition between high and low reaction temperatures for any given system.

The difference in yields due to reaction temperature led to the development of a simple ligation model that allowed us to determine the key factors that could influence ligation yields. This model will also be useful for others in the design of their ligation system. Despite optimization of our particular DNA system, there was still sequence dependence at the ligating nick, similar to findings by Damha and coworkers (44). Thus, highlighting yet another factor necessary for designing a non-enzymatic ligating system, challenges that are absent in the use of enzymes for ligation.

Having developed a framework with which we could understand non-enzymatic ligation within the DNA systems, the latter part of the thesis focused on investigating the

efficacy of ligation in other systems. The position of the phosphate for DNA and RNA ligating systems were investigated, in addition to the effect of the sugar on ligation yields. It was found that for DNA systems, a 3´-phosphate was better than a 5´-phosphate possibly due to the presence of a primary vs secondary alcohol. On the converse, for RNA ligation, a 5´-phosphate terminus was better than a 3´-phosphate due to the formation of a cyclic phosphate in the latter.

Given the role of cyclic phosphates in prebiotic chemistry, the last chapter of this Thesis, used a ribosomal template to investigate the finding by others that ligation in loops could be better than ligation in an adjacent end-end ligating system (24,51). These cyclic phosphate ligation studies revealed the large role of sequence dependence within the ligating system. In particular, the ease of incorporation of certain mismatches at the ligation nick demonstrates one possible pathway to the prebiotic formation of internal and terminal loops present in known ribosomes and ribozymes (96,111,112,130,131).

A major limitation to high yields in cyclic phosphate ligation, which the systematic nature of this work revealed was hydrolysis of the substrates. Due to the long reaction times (several days) required for ligation, much of the substrate is lost to hydrolysis, thereby limiting the number of substrates available for ligation. It is possible that this was a major factor to the low yields often experienced in works by others, especially due to the presence of divalent ions required for ligation (27,51,107).

Overall, the discoveries in this Thesis led to the development of a framework which laid out the most important factors responsible for attaining high yields in both a DNA and RNA ligating system.

6.2 Future directions

Despite the progress presented in this dissertation, there are still several unanswered questions that remain in the field of non-enzymatic ligation, and especially in the development of an enzyme free replication system. An open question that remains is the role of sequences in the selection of a functional biopolymer. Results from both the DNA and RNA ligating systems revealed that the identity of sequences at the ligation nick can have a large impact on yields. These results build on the seminal works by Ellington and coworkers (73,132) which addressed the fidelity of DNA non-enzymatic ligation.

To further the DNA studies, it would be interesting to design a competition experiment that involves different lengths of substrates, in order to decipher which substrates could be selected in a prebiotic pool. At colder temperatures, where most substrates are bound, how does selectivity come into play? Another question that can be addressed alongside DNA selectivity is the availability of activating agents. Similar to work by Edeleva and coworkers (49) in which substrates were fed into the system, will the addition of EDC periodically into the system lead to an increase in ligation yields? Addressing questions such as those posed above would push the field towards the development of a replication cycle which involves heating and cooling.

With regards to RNA cyclic phosphate ligation, the studies performed in this thesis were built around a limited set of sequences in order to finely address ligation efficacy. To further this work and make it accessible to larger members of the prebiotic field, a more robust study should be developed that begins with random sequences at the ligation nick

and undergoes evolution to select for the best ligation sequence. Such a sequence selection study will be akin to that performed by Mutschler and coworkers (82), but would be different given that the selection will be based solely on sequences around the ligation nick. Perhaps, a 2-step selection can be performed in which the nick substrates are first varied, followed by the varying of the pairing partners on the template strand, in order to study the interdependence of the two. By performing a wider search for optimum ligating sequence, the field would learn nature's rules concerning which sequences and junctions within an RNA molecule would be most likely to have evolved using a cyclic phosphate intermediate. It is also possible that such an evolution study will lead to the selection of sequences that yield higher products than were observed in this thesis and might exclusively possess the often elusive canonical 3´-5´ linkages.

Another major area of research is the recyclability of substrates. Given the prevalent role of hydrolysis in limiting product yield, studies should be designed that investigate the ability of the hydrolyzed cyclic phosphate to be rescued and reused in ligation. Similar to work from the Mutschler lab with a ribozyme system (33), prebiotic plausible activating agents such as DAP, which could have been present on the prebiotic earth can be added into the reaction mix under a prebiotically plausible cycle, to determine whether all of the limiting reagent will be consumed and maximum ligation yield reached.

Lastly, the use of non-aqueous solvents for enzyme free replication has been reported in recent years (11,12), but it's full potential has not yet been achieved due to the use of enzymatic ligation. In order to address this shortcoming, and to expand the use chemical ligation beyond aqueous solvents, the efficacy of cyclic phosphate ligation in non-aqueous solvents needs to be investigated. There are several potential challenges that

could arise from this undertaking. One of the first questions that need to be addressed is the stability of the cyclic phosphate substrate in these non-aqueous solvents. The presence of alcohols and denaturing agents in several published non-aqueous solvents (133,134) might lead to faster cyclic phosphate hydrolysis compared to aqueous solvents. The stability of the substrates also needs to be investigated with regards to prebiotic cycles. Once the role of cycles is understood, these cycles can then be used to demonstrate successive increases in replication yields. The optimization of ligation in non-aqueous solvents will provide a path to the development of a true enzyme free replicating system.

In conclusion, this thesis has provided a systematic lens through which non-enzymatic ligation can now be understood. For prebiotic chemists, the limitations of cyclic phosphate ligation were addressed, highlighting in particular, the hurdles to enzyme free RNA replication. Given the difficulties present with RNA ligation, and replication as whole, it is most likely the case that proto-RNA molecules such as can be assembled by hexad assemblies or peptide backbones were the first prebiotic polymers on the early earth (135,136), and not RNA. For others interested in the development of nucleic acid nanotechnology, a useful framework was developed that could facilitate the development of covalent structures for DNA nanotechnology (137) and molecular sensing probes (138,139).

APPENDIX A. MATLAB FILES FOR LIGATION MODEL AND CONTOUR PLOT

For the 9-mer.

Side product decay rate

```
% This file contains two functions. The first function is the function
% called in the equation solver for optimization. The second function is
to
% define the differential equations
function Optm = sideprodeqn_mod_0501(k)
% k(1) = kc, and k(2) = ks as described in main function
% all units are in micromolar
%% ode model for overall ligation
function [ dalldt ] = Ligratfun_0608_CIP(t,cn)
%% initializing odes. S = side products, I = hybridized inermediates, P
= products
% H = hairpin concentration, and O = substrate concentration, E = EDC
 O = cn(1);
 E = cn(2);
 % full equation
```

```
 dOdt = -O*E*k(1);

 dEdt = -E*O*k(1) -E*k(2);

dalldt = [dOdt; dEdt];

end

%% Definining initial values and experimental yields

t = [0 0.5 2 4 10 24];%time in hrs

j =3; %change for temps i.e. 1 = 4C, 2 = 25C, and 3 = 37C

% yields of the 9-mer reactions. 1st row = 4C, 2nd row = 25C, and third
row

% = 37C

ninemer =[100 100 100

98.5 96 88

98  91  69

97  84  58

93  71  48

87  56  49

]*20e-2

; %converts side products to concentration

 % initial conditions for the ode solver
```

```
% S0 = 0; %initial concentration of side products

E0 = 250; % initial concentration of EDC (in mM)

O0 = 20; %initial concentration of substrate (in micromolar)

% O0 = ninemer(1,j); % Experimental initial concentration of 9-mer'

Oexp = ninemer(:,j); % Calling experimental yields

%% Numerical Intgration to solve the odes

opts = odeset('RelTol',1e-5,'AbsTol',1e-5);

[t,cn]= ode23s(@ Ligratfun_0608_CIP,t,[O0 E0],opts);

Oth = cn(:,1); % predicted substrate loss

T = table(cn) % summary of predicted side reactants, intermediate,ligated
products, hairpin, and substrate concentration

%% Optimization/Minimization of the predicted and experimental values

Optm = sum ((Oth-Oexp).^2)

plot(t,Oth,'-*',t,Oexp,'o')

end

% This file solves the equation for finding the side product constant

clc

clear all

close all
```

```
%% Overall goal: Find the chemical ligation constant kc

% and side product reaction constants ks

%% solving the overall ode 9 mer

k = [0 0]; %initial guess for rate constants. k(1) == kc, k(2) == ks

LB = [ ]; % lower bounds on k

UB = [1 1]; % upper bounds on k

% Function to minimize the experimental and predicted yields using

% equations defined in file named Ligationconstant_eqn

% options = optimset('TolCol', 1e-20);

[kopt,fval] = fmincon (@sideprodeqn_mod_0501v3,k,[],[],[],[],LB,UB,[])
```

Ligation reaction optimization to find rate constant

```
% This file contains two functions. The first function is the function

% called in the equation solver for optimization. The second function is

to

% define the differential equations

function Optm = Ligationconstant_eqn_1pt5x_0309(k)

% k = ligation reaction

% all units are in micromolar

%% ode model for overall ligation
```

```
function [ dalldt ] = Ligratfun_0608_CIP(t,cn)

 global Keq %calling the equilibrium constant

%% initializing odes. S = side products, I = hybridized inermediates, P
= products

% H = hairpin concentration, and O = substrate concentration

 I = cn(1);

 P = cn(2);

 H = cn(3);

 O = cn(4);

 E = cn(5);

 %% rate constants. kf = forward rate constant, Keq = equilibrium constant

 % For all rate constants the values are always ordered 4C, 25C, and 37C

%

 kf = [1e4];% units 1/uM.hr

kh = [0.036 0.067 0.24];% hydrolysis rate constant hr-1

ks = [3.5e-5 0.0002 8e-4];% side reaction rate constant hr-1.mM-1

 dIdt = kf*(H*O - ((1/Keq(j))*I)) - k*I*E;

 dPdt = k*I*E;

 dHdt = kf*(-H*O + ((1/Keq(j))*I)) - ks(j)*H*E;
```

```
 dOdt = kf*(-H*O + ((1/Keq(j))*I)) - ks(j)*O*E;

 dEdt = -k*I*E - ks(j)*H*E - ks(j)*O*E -E*kh(j);



 dalldt = [dIdt; dPdt; dHdt; dOdt; dEdt];

end

%% Definining initial values and experimental yields

t = [0 0.5 2 4 6 10 24];%time in hrs

j = 2; %change for temps i.e. 1 = 4C, 2 = 25C, and 3 = 37C

% yields of the 9-mer reactions. 1st row = 4C, 2nd row = 25C, and third
row

% = 37C

% for a 10x rxn

ninemer =[0 0 0

    19.21556924 34.48291562 52.3652643

36.0658221  63.2110746  82.76368563

48.91502769 80.63389711 88.58825526

56.04074855 86.44799309 89.36796167

68.29594527 90.54507854 91.1213863
```

```
84.86050128 94.97187274 90.77848076]*1.3e-2; %converts ligated products
to concentration

% initial conditions for the ode solver

I0 = 0;

E0 = 250;

P0 = ninemer(1,j); % Experimental initial concentration of 9-mer'

% ex = 1.25; %excess of oligos

ex = 10; %excess of oligos

H0 = 1.3; %initial conc on hairpin in (micromolar)

O0 = ex*H0; %initial concentration of substrate

Pexp = ninemer(:,j); % Calling experimental yields

%% Numerical Intgration to solve the odes

opts = odeset('RelTol',1e-7,'AbsTol',1e-7);

[t,cn]= ode23s(@ Ligratfun_0608_CIP,t,[I0 P0 H0 O0 E0],opts);

Pth = cn(:,2); % predicted products

T_sum = table(cn) % summary of predicted side reactants,
intermediate,ligated products, hairpin, and substrate concentration

%% Optimization/Minimization of the predicted and experimental values

Optm = sum ((Pth-Pexp).^2)

plot(t,Pth,'-*',t,Pexp,'o')
```

```
end



% This file finds the equilibrium constant (Keq) and then optimizes
% the experimental and predicted values for
% the chemical ligation rate constant k
% for the 9-mer 25C reaction
% using the differntial equations described in the Supplementary
information
clc
clear all
close all
global Keq
%% Overall goal: Find the chemical ligation constant kc
% and side product reaction constants ks
%% Finding Equilibrium constant for the 9mer
deltaH_9 = -261668; %J/mol Enthalpy of the 9-mer substrate
Tm9= [35]+273; % Melting temperature (Tm) of the 9-mer in Kelvin for a
1.5x rxn
```

```
% Tm9= [43]+273; % Melting temperature (Tm) of the 9-mer in Kelvin for a
10x rxn

T = [4 25 37] + 273;% Reaction temp in Kelvin

R = 8.314; %J.K-1mol-1



H = 1.3 *10^-6; %uM Hairpin concentration

e = [1.5];

% e = 10;

x = 0.5;

KeqTm = (x/(H*(1-x)*(e-x))); % Equilibrium constant at the Tm

Keq = (KeqTm./(exp((deltaH_9/R).*((1./T)-(1/Tm9))))))./1e6; % Equilibrium
constant at reaction T converted to micromolar units

k = [1e-4]; %initial guess for rate constants. k(1) == kc, k(2) == ks

LB = [0]; % lower bounds on k

UB = []; % upper bounds on k

% Function to minimize the experimental and predicted yields using

% equations defined in file named Ligationconstant_eqn

[kopt,fval]                              =                              fmincon
(@Ligationconstant_eqn_1pt5x_0309,k,[],[],[],[],LB,UB,[])
```

For the 5-mer.

Contour plot

```
%% function file for contour plot

%% odes for finding S,I, and P

function [ dalldt ] = Contourplot_eqn_wHmod(t,cn,kc,ks,Keq,kf,kh)

 I = cn(1);

 P = cn(2);

 H = cn(3);

 O = cn(4);

 E = cn(5);

 % rate constants. kf = forward rate constant, kc = chemical reaction
rate

 % constant

 % Keq = equilibrium constant, ks = side reaction constant

 dIdt = kf*(H*O - ((1/Keq)*I)) - kc*I*E;

 dPdt = kc*I*E;

 dHdt = kf*(-H*O + ((1/Keq)*I)) - ks(1)*H*E;

 dOdt = kf*(-H*O + ((1/Keq)*I)) - ks(2)*O*E;

 dEdt = -(kc*I*E) - (ks(1)*H*E) - (ks(2)*O*E) -E*kh;
```

```
 dalldt = [dIdt; dPdt; dHdt; dOdt; dEdt];
end
% This file constructs the contour plot using the predicted rate constants
% for the 5-mer at 25C reaction


clc
clear all
close all
%%
% initial conditions for side reactions and intermediate hybrid
I0 = 0;
P0 = 0;
E0 = 250;
%% rate constants. kf = forward rate constant, kc = chemical reaction
rate
 % constant Keq = equilibrium constant, ks = side reaction constant
 kf =  1e4; %forward reaction rate constant units 1/uM.mM.hr
 kc = 0.0021;% chemical  reaction constant mM -1hr-1
 ks = [5.14e-5 2.2e-6];% side reaction rate constant hr-1.mM-1
```

```
 kh = 0.067;% hydrolysis rate constant hr-1

time = [0:0.1:24];

 %% values to make contour

Keq = logspace(-3,0.5,30); %Range of Equilibrium constant

m = length (Keq);

H0 = 1.3; %initial conc on hairpin (micromolar)

Subconc = logspace(0,2.9,30); %substrate concentration

ex = Subconc/H0; % susbtrate:hairpin excess ratio

n = length (ex);

O0 = ex.*H0;%initial concentration of substrate

% initialize matrix to store values of products

P = zeros(m,n);

% loop to calculate product yield using the odes and storing the 24 hr
yield in P matrix

for a = 1:m

    for b = 1:n

%% Numerical Intgration to solve the odes

opts = odeset('RelTol',1e-7,'AbsTol',1e-7);
```

```
[t,cn]=                          ode23s(@                          (t,cn)
Contourplot_eqn_wHmod(t,cn,kc,ks,Keq(a),kf,kh),time,[I0       P0       H0
O0(b),E0],opts);

P(b,a) = (cn(end,2)/H0).*101;

    end

end

% commands for plotting the contour

pc = contourf(Keq.*H0,ex,P,'ShowText','on');

ax = gca;

ax.YScale = 'log';

ax.XScale = 'log';

h = colorbar;

title(h,'Yield','FontSize',15,'FontWeight','bold','LineWidth',2')

xlabel('Keq*[Hairpin]','FontSize',15,'FontWeight','bold','LineWidth',2)

ylabel                                                              ('
[Substrate]/[Hairpin]','FontSize',15,'FontWeight','bold','LineWidth',2'
)

%ax.XAxisLocation = 'origin';

ax.Box = 'off';

% ax.Layer = 'top';
```

```
set(gca,'TickDir','out','linewidth',2,'fontweight','bold','fontsize',15
);
```

APPENDIX B. ADDITIONAL DATA AND FIGURES

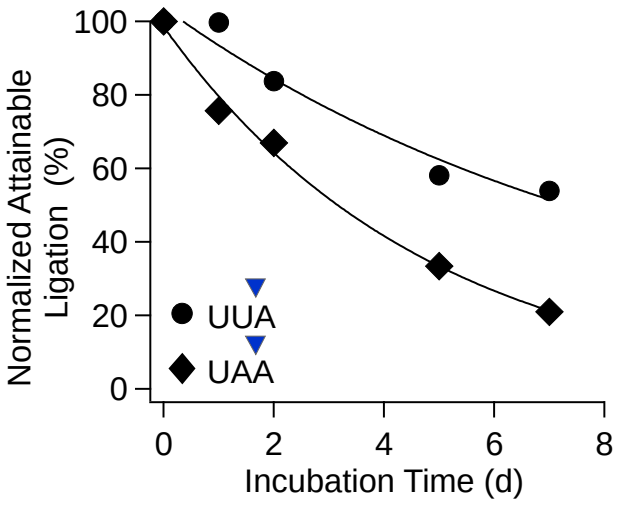

Figure 46. Cyclic phosphate hydrolysis of two different substrates with a U and an A base-pair. Similar to studies conducted in Section 5.5.

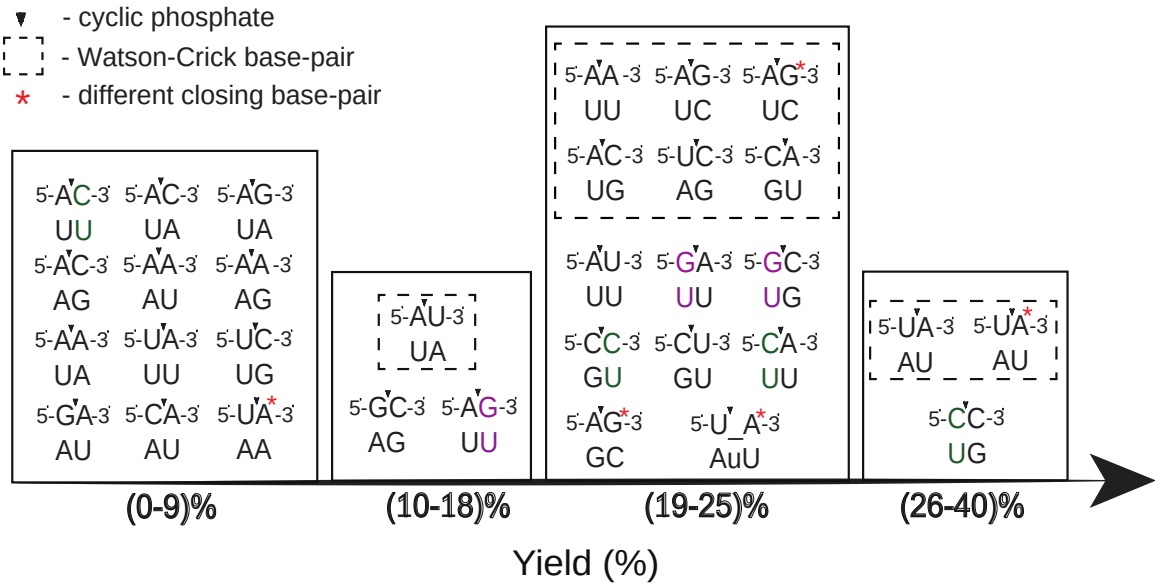

Figure 47. Yield diagram showing summary of cyclic phosphate ligation results studied in CHAPTER 5.

REFERENCES

1. Peacocke, A. and Drysdale, R. (1965), *The Molecular Basis of Heredity*. Springer, pp. 11-20.

2. Gilbert, W. (1986) Origin of life: The RNA world. *Nature*, **319**, 618-618.

3. Orgel, L.E. (1994) The origin of Life on Earth. *Scientific American*, **271**, 76-83.

4. Orgel, L.E. (1998) The origin of life—a review of facts and speculations. *Trends in biochemical sciences*, **23**, 491-495.

5. Pace, N.R. and Marsh, T.L. (1985) RNA catalysis and the origin of life. *Origins of Life and Evolution of the Biosphere*, **16**, 97-116.

6. Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature*, **418**, 214-221.

7. Joyce, G.F. and Orgel, L.E. (1993) Prospects for understanding the origin of the RNA world. *Cold Spring Harbor Monograph Series*, **24**, 1-1.

8. Szostak, J.W. (2017) The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angew Chem Int Ed Engl*, **56**, 11037-11043.

9. Monnard, P.-A. (2016) Taming prebiotic chemistry: the role of heterogeneous and interfacial catalysis in the emergence of a prebiotic catalytic/information polymer system. *Life*, **6**, 40.

10. Szostak, J.W. (2012) The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry*, **3**, 2.

11. He, C., Gallego, I., Laughlin, B., Grover, M.A. and Hud, N.V. (2017) A viscous solvent enables information transfer from gene-length nucleic acids in a model prebiotic replication cycle. *Nat Chem*, **9**, 318-324.

12. He, C., Lozoya-Colinas, A., Gallego, I., Grover, M.A. and Hud, N.V. (2019) Solvent viscosity facilitates replication and ribozyme catalysis from an RNA duplex in a model prebiotic process. *Nucleic Acids Res*, **47**, 6569-6577.

13. Giurgiu, C., Li, L., O'Flaherty, D.K., Tam, C.P. and Szostak, J.W. (2017) A Mechanistic Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension. *J Am Chem Soc*, **139**, 16741-16747.

14. Zhang, W., Tam, C.P., Walton, T., Fahrenbach, A.C., Birrane, G. and Szostak, J.W. (2017) Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc Natl Acad Sci U S A*, **114**, 7659-7664.

15. Zhou, L., O'Flaherty, D.K. and Szostak, J.W. (2020) Template-directed copying of RNA by non-enzymatic ligation. *Angewandte Chemie*, **132**, 15812-15817.

16. Zielinski, W.S. and Orgel, L.E. (1987) Autocatalytic synthesis of a tetranucleotide analogue. *Nature*, **327**, 346-347.

17. von Kiedrowski, G. (1986) A self-replicating hexadeoxynucleotide. *Angewandte Chemie International Edition in English*, **25**, 932-935.

18. Deck, C., Jauker, M. and Richert, C. (2011) Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat Chem*, **3**, 603-608.

19. Jauker, M., Griesser, H. and Richert, C. (2015) Copying of RNA Sequences without Pre-Activation. *Angew Chem Int Ed Engl*, **54**, 14559-14563.

20. Kervio, E., Sosson, M. and Richert, C. (2016) The effect of leaving groups on binding and reactivity in enzyme-free copying of DNA and RNA. *Nucleic Acids Res*, **44**, 5504-5514.

21. Sosson, M., Pfeffer, D. and Richert, C. (2019) Enzyme-free ligation of dimers and trimers to RNA primers. *Nucleic Acids Res*.

22. Powner, M.W., Gerland, B. and Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, 239-242.

23. Verlander, M.S., Lohrmann, R. and Orgel, L.E. (1973) Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J Mol Evol*, **2**, 303-316.

24. Nechaev, S.Y., Lutay, A.V., Vlassov, V.V. and Zenkova, M.A. (2009) Non-enzymatic template-directed recombination of RNAs. *Int J Mol Sci*, **10**, 1788-1807.

25. Hertel, K.J., Herschlag, D. and Uhlenbeck, O.C. (1994) A kinetic and thermodynamic framework for the hammerhead ribozyme reaction. *Biochemistry*, **33**, 3374-3385.

26. Lutay, A.V., Chernolovskaya, E.L., Zenkova, M.A. and Vlasov, V.V. (2005) Nonenzymatic template-dependent ligation of 2',3'-cyclic phosphate-containing oligonucleotides catalyzed by metal ions. *Dokl Biochem Biophys*, **401**, 163-166.

27. Lutay, A.V., Chernolovskaya, E.L., Zenkova, M.A. and Vlassov, V.V. (2006) The nonenzymatic template-directed ligation of oligonucleotides. *Biogeosciences Discussions*, **3**, 1-21.

28. Crowe, M.A. and Sutherland, J.D. (2006) Reaction of Cytidine Nucleotides with Cyanoacetylene: Support for the Intermediacy of Nucleoside-2′, 3′-cyclic Phosphates in the Prebiotic Synthesis of RNA. *ChemBioChem*, **7**, 951-956.

29. Lohrmann, R. and Orgel, L. (1968) Prebiotic synthesis: phosphorylation in aqueous solution. *Science*, **161**, 64-66.

30. Lohrmann, R. and Orgel, L. (1971) Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science*, **171**, 490-494.

31. Gibard, C., Bhowmik, S., Karki, M., Kim, E.K. and Krishnamurthy, R. (2018) Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat Chem*, **10**, 212-217.

32. Gibard, C., Gorrell, I.B., Jiménez, E.I., Kee, T.P., Pasek, M.A. and Krishnamurthy, R. (2019) Geochemical sources and availability of amidophosphates on the early Earth. *Angewandte Chemie*, **131**, 8235-8239.

33. Song, E.Y., Jiménez, E.I., Lin, H., Le Vay, K., Krishnamurthy, R. and Mutschler, H. (2020) Prebiotically plausible RNA activation compatible with ribozyme-catalyzed ligation. *Angewandte Chemie*.

34. Seeman, N.C. (2010) Nanomaterials based on DNA. *Annual review of biochemistry*, **79**, 65-87.

35. Rothemund, P.W. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297-302.

36. Yaradoddi, J.S., Kontro, M.H., Ganachari, S.V., Sulochana, M.B., Agsar, D., Tapaskar, R.P. and Shettar, A.S. (2019) DNA Nanotechnology. *Handbook of Ecomaterials. Springer, Cham.*, 3561-3572.

37. Kumar, R., El-Sagheer, A., Tumpane, J., Lincoln, P., Wilhelmsson, L.M. and Brown, T. (2007) Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *Journal of the American Chemical Society*, **129**, 6859-6864.

38. El-Sagheer, A.H. and Brown, T. (2012) Click nucleic acid ligation: applications in biology and nanotechnology. *Accounts of chemical research*, **45**, 1258-1267.

39. Kanan, M.W., Rozenman, M.M., Sakurai, K., Snyder, T.M. and Liu, D.R. (2004) Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature*, **431**, 545-549.

40. Gartner, Z.J., Grubina, R., Calderone, C.T. and Liu, D.R. (2003) Two enabling architectures for DNA-templated organic synthesis. *Angewandte Chemie International Edition*, **42**, 1370-1375.

41. Maruyama, H., Oikawa, R., Hayakawa, M., Takamori, S., Kimura, Y., Abe, N., Tsuji, G., Matsuda, A., Shuto, S. and Ito, Y. (2017) Chemical ligation of oligonucleotides using an electrophilic phosphorothioester. *Nucleic acids research*, **45**, 7042-7048.

42. Dolinnaya, N.G., Sokolova, N.I., Ashirbekova, D.T. and Shabarova, Z.A. (1991) The use of BrCN for assembling modified DNA duplexes and DNA-RNA hybrids; comparison with water-soluble carbodiimide. *Nucleic Acids Research*, **19**.

43. Sokolova, N.I., Ashirbekova, D.T., Dolinnaya, N.G. and Shabarova, Z.A. (1988) Chemical reactions within DNA duplexes. Cyanogen bromide as an effective oligodeoxyribonucleotide coupling agent. *FEBS Lett*, **232**, 153-155.

44. Carriero, S. and Damha, M.J. (2002) Synthesis of Lariat-DNA via the Chemical Ligation of a Dumbbell Complex. *Organic Letters*, **5**.

45. Harada, K. and Orgel, L.E. (1994) In Vitro Selection of Optimal DNA Substrates for Ligation by a Water-Soluble Carbodiimide. *J Mol Evol*, **38**.

46. Kramer, M. and Richert, C. (2017) Enzyme-Free Ligation of 5'-Phosphorylated Oligodeoxynucleotides in a DNA Nanostructure. *Chem Biodivers*, **14**.

47. Xu, Y. and Kool, E.T. (1999) High sequence fidelity in a non-enzymatic DNA autoligation reaction. *Nucleic Acids Res*, **27**, 875-881.

48. Dolinnaya, N.G., Sokolova, N.I., Gryaznova, O.I. and Shabarova, Z.A. (1988) Site-directed modification of DNA duplexes by chemical ligation. *Nucleic Acids Res*, **16**, 3721-3738.

49. Edeleva, E., Salditt, A., Stamp, J., Schwintek, P., Boekhoven, J. and Braun, D. (2019) Continuous nonenzymatic cross-replication of DNA strands with in situ activated DNA oligonucleotides. *Chemical Science*, **10**, 5807-5814.

50. Lutay, A.V., Zenkova, M.A. and Vlassov, V.V. (2007) Nonenzymatic recombination of RNA: possible mechanism for the formation of novel sequences. *Chem Biodivers*, **4**, 762-767.

51. Staroseletz, Y., Nechaev, S., Bichenkova, E., Bryce, R.A., Watson, C., Vlassov, V. and Zenkova, M. (2018) Non-enzymatic recombination of RNA: Ligation in loops. *Biochim Biophys Acta Gen Subj*, **1862**, 705-725.

52. Nichols, N.M., Tabor, S. and McReynolds, L.A. (2008) RNA ligases. *Curr Protoc Mol Biol*, **Chapter 3**, Unit3 15.

53. Sgaramella, V. and Khorana, H.G. (1992) A further study of the T4 ligase-catalyzed joining of DNA at base-paired ends. 1972. *Biotechnology*, **24**, 28-37.

54. Sugino, A., Goodman, H.M., Heyneker, H.L., Shine, J., Boyer, H.W. and Cozzarelli, N.R. (1977) Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base-paired ends. *J Biol Chem*, **252**, 3987-3994.

55. Sugino, A., Snoper, T.J. and Cozzarelli, N.R. (1977) Bacteriophage T4 RNA ligase. Reaction intermediates and interaction of substrates. *J Biol Chem*, **252**, 1732-1738.

56. Tsytovich, A.V., Dolinnaia, N.G. and Shabarova, Z.A. (1988) T4-DNA ligase: substrate properties of synthetic DNA-duplexes with structural anomalies. *Mol Biol (Mosk)*, **22**, 690-699.

57. Monnard, P.A. and Szostak, J.W. (2008) Metal-ion catalyzed polymerization in the eutectic phase in water-ice: a possible approach to template-directed RNA polymerization. *J Inorg Biochem*, **102**, 1104-1111.

58. Prywes, N., Blain, J.C., Del Frate, F. and Szostak, J.W. (2016) Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *Elife*, **5**.

59. Walton, T. and Szostak, J.W. (2017) A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5'-phosphoro-2-aminoimidazolide. *Biochemistry*, **56**, 5739-5747.

60. Ferris, J.P., Hill, A.R., Jr., Liu, R. and Orgel, L.E. (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, **381**, 59-61.

61. Novère, L. (2001), Bioinformatics, Vol. 17, pp. 1226-1227.

62. SantaLucia, J., Jr. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct*, **33**, 415-440.

63. Horowitz, E.D., Engelhart, A.E., Chen, M.C., Quarles, K.A., Smith, M.W., Lynn, D.G. and Hud, N.V. (2010) Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. *Proc Natl Acad Sci U S A*, **107**, 5288-5293.

64. Todisco, M., Fraccia, T.P., Smith, G.P., Corno, A., Bethge, L., Klussmann, S., Paraboschi, E.M., Asselta, R., Colombo, D., Zanchetta, G. *et al.* (2018) Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly. *ACS Nano*, **12**.

65. Nwokeoji, A.O., Kilby, P.M., Portwood, D.E. and Dickman, M.J. (2017) Accurate Quantification of Nucleic Acids Using Hypochromicity Measurements in Conjunction with UV Spectrophotometry. *Anal Chem*, **89**, 13567-13574.

66. Rohatgi, R., Bartel, D.P. and Szostak, J.W. (1996) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J Am Chem Soc*, **118**, 3332-3339.

67. SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A*, **95**.

68. Mergny, J.L. and Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515-537.

69. Cavaluzzi, M.J. and Borer, P.N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res*, **32**, e13.

70. Gilham, P. (1962) An addition reaction specific for uridine and guanosine nucleotides and its application to the modification of ribonuclease action. *Journal of the American Chemical Society*, **84**, 687-688.

71. Eichinger, B. (2000) Cyclization in reversible and irreversible step-growth polymerizations. *Computational and Theoretical Polymer Science*, **10**, 83-88.

72. Kawamura, K. and Okamoto, F. (2001) Cyclization and dimerization of hexanucleotides containing guanine and cytosine with water-soluble carbodiimide. *Viva Origino*, **29**, 162.

73. James, K.D. and Ellington, A.D. (1997) Surprising fidelity of template-directed chemical ligation of oligonucleotides. *Chem Biol*, **4**, 595-605.

74. Tena-Solsona, M., Rieß, B., Grötsch, R.K., Löhrer, F.C., Wanzke, C., Käsdorf, B., Bausch, A.R., Müller-Buschbaum, P., Lieleg, O. and Boekhoven, J. (2017) Non-equilibrium dissipative supramolecular materials with a tunable lifetime. *Nature communications*, **8**, 1-8.

75. Wrobel, N., Schinkinger, M. and Mirsky, V.M. (2002) A novel ultraviolet assay for testing side reactions of carbodiimides. *Anal Biochem*, **305**, 135-138.

76. Norberg, J. and Nilsson, L. (1995) Stacking free energy profiles for all 16 natural ribodinucleoside monophosphates in aqueous solution. *Journal of the American Chemical Society*, **117**, 10832-10840.

77. Norberg, J. and Nilsson, L. (1998) Solvent influence on base stacking. *Biophys J*, **74**, 394-402.

78. Šponer, J., Šponer, J.E., Mládek, A., Jurečka, P., Banáš, P. and Otyepka, M. (2013) Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*, **99**, 978-988.

79. Ashley, G.W. and Kushlan, D.M. (1991) Chemical synthesis of oligodeoxynucleotide dumbbells. *Biochemistry*, **30**, 2927-2933.

80. Yakovchuk, P., Protozanova, E. and Frank-Kamenetskii, M.D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research*, **34**, 564-574.

81. Ho, N.W. and Gilham, P.T. (1967) The reversible chemical modification of uracil, thymine, and guanine nucleotides and the modification of the action of ribonuclease on ribonucleic acid. *Biochemistry*, **6**, 3632-3639.

82. Mutschler, H., Taylor, A.I., Porebski, B.T., Lightowlers, A., Houlihan, G., Abramov, M., Herdewijn, P. and Holliger, P. (2018) Random-sequence genetic oligomer pools display an innate potential for ligation and recombination. *Elife*, **7**.

83. Smail, B.A., Clifton, B.E., Mizuuchi, R. and Lehman, N. (2019) Spontaneous advent of genetic diversity in RNA populations through multiple recombination mechanisms. *RNA*, **25**, 453-464.

84. James, K.D. and Ellington, A.D. (1999) The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Orig Life Evol Biosph*, **29**, 375-390.

85. Holbrook, S.R., Cheong, C., Tinoco, I. and Kim, S.-H. (1991) Crystal structure of an RNA double helix incorporating a track of non-Watson–Crick base pairs. *Nature*, **353**, 579-581.

86. Dock-Bregeon, A., Chevrier, B., Podjarny, A., Moras, D., DeBear, J., Gough, G., Gilham, P. and Johnson, J. (1988) High resolution structure of the RNA duplex [U (UA) 6A] 2. *Nature*, **335**, 375-378.

87. Nelson, H.C., Finch, J.T., Luisi, B.F. and Klug, A. (1987) The structure of an oligo (dA)· oligo (dT) tract and its biological implications. *Nature*, **330**, 221-226.

88. Hud, N.V. and Polak, M. (2001) DNA–cation interactions: the major and minor grooves are flexible ionophores. *Current opinion in structural biology*, **11**, 293-301.

89. Lie, L., Biliya, S., Vannberg, F. and Wartell, R.M. (2016) Ligation of RNA Oligomers by the Schistosoma mansoni Hammerhead Ribozyme in Frozen Solution. *J Mol Evol*, **82**, 81-92.

90. Lehman, N. (2008) A Recombination-Based Model for the Origin and Early Evolution of Genetic Information. *Chem Biodivers*, **5**, 1707-1717.

91. Chetverin, A.B., Chetverina, H.V., Demidenko, A.A. and Ugarov, V.I. (1997) Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell*, **88**, 503-513.

92. Chetverina, H.V., Demidenko, A.A., Ugarov, V.I. and Chetverin, A.B. (1999) Spontaneous rearrangements in RNA sequences. *FEBS letters*, **450**, 89-94.

93. Verlander, M. and Orgel, L. (1974) Analysis of high molecular weight material from the polymerization of adenosine cyclic 2′, 3′-phosphate. *Journal of molecular evolution*, **3**, 115-120.

94. Usher, D. and McHale, A. (1976) Nonenzymic joining of oligoadenylates on a polyuridylic acid template. *Science*, **192**, 53-54.

95. Mutschler, H. and Holliger, P. (2014) Non-canonical 3'-5' extension of RNA with prebiotically plausible ribonucleoside 2',3'-cyclic phosphates. *J Am Chem Soc*, **136**, 5193-5196.

96. Mutschler, H., Wochner, A. and Holliger, P. (2015) Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nat Chem*, **7**, 502-508.

97. Attwater, J., Wochner, A., Pinheiro, V.B., Coulson, A. and Holliger, P. (2010) Ice as a protocellular medium for RNA replication. *Nature Communications*, **1**, 1-9.

98. Kazakov, S.A., Balatskaya, S.V. and Johnston, B.H. (2006) Ligation of the hairpin ribozyme in cis induced by freezing and dehydration. *Rna*, **12**, 446-456.

99. Berk, C., Wang, Y., Laski, A., Tsagkris, S. and Hall, J. (2020) Ligation of 2′, 3′-cyclic phosphate RNAs for the identification of microRNA binding sites. *FEBS letters*.

100. Chakravarty, A.K. and Shuman, S. (2012) The sequential 2′, 3′-cyclic phosphodiesterase and 3′-phosphate/5′-OH ligation steps of the RtcB RNA splicing pathway are GTP-dependent. *Nucleic acids research*, **40**, 8558-8567.

101. Tanaka, N., Chakravarty, A.K., Maughan, B. and Shuman, S. (2011) Novel mechanism of RNA repair by RtcB via sequential 2′, 3′-cyclic phosphodiesterase and 3′-Phosphate/5′-hydroxyl ligation reactions. *Journal of Biological Chemistry*, **286**, 43134-43143.

102. Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K., Trowitzsch, S., Will, C.L., Lührmann, R. and Söll, D. (2011) HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science*, **331**, 760-764.

103. Eftink, M.R. and Biltonen, R.L. (1983) Energetics of ribonuclease A catalysis. 2. Nonenzymatic hydrolysis of cytidine cyclic 2', 3'-phosphate. *Biochemistry*, **22**, 5134-5140.

104. Bowler, F.R., Chan, C.K., Duffy, C.D., Gerland, B., Islam, S., Powner, M.W., Sutherland, J.D. and Xu, J. (2013) Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat Chem*, **5**, 383-389.

105. Hüsken, D., Goodall, G., Blommers, M.J., Jahnke, W., Hall, J., Häner, R. and Moser, H.E. (1996) Creating RNA bulges: cleavage of RNA in RNA/DNA duplexes by metal ion catalysis. *Biochemistry*, **35**, 16591-16600.

106. Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol*, **39**, 99-123.

107. Lutay, A.V., Grigoriev, I.V., Zenkova, M.A., Chernolovskaya, E.L. and Vlassov, V.V. (2007) Nonenzymatic recombination of RNA by means of transesterification. *Russian Chemical Bulletin*, **56**, 2499-2505.

108. Shankar, N., Kennedy, S.D., Chen, G., Krugh, T.R. and Turner, D.H. (2006) The NMR structure of an internal loop from 23S ribosomal RNA differs from its structure in crystals of 50S ribosomal subunits. *Biochemistry*, **45**, 11776-11789.

109. Lee, J.C., Gutell, R.R. and Russell, R. (2006) The UAA/GAN internal loop motif: a new RNA structural element that forms a cross-strand AAA stack and long-range tertiary interactions. *Journal of molecular biology*, **360**, 978-988.

110. Réblová, K., Střelcová, Z., Kulhánek, P., Beššeová, I., Mathews, D.H., Van Nostrand, K., Yildirim, I., Turner, D.H. and Sponer, J. (2010) An RNA molecular switch: Intrinsic flexibility of 23S rRNA helices 40 and 68 5′-UAA/5′-GAN internal loops studied by molecular dynamics methods. *Journal of chemical theory and computation*, **6**, 910-929.

111. Petrov, A.S., Bernier, C.R., Hsiao, C., Norris, A.M., Kovacs, N.A., Waterbury, C.C., Stepanov, V.G., Harvey, S.C., Fox, G.E. and Wartell, R.M. (2014) Evolution of the ribosome at atomic resolution. *Proceedings of the National Academy of Sciences*, **111**, 10251-10256.

112. Petrov, A.S., Gulen, B., Norris, A.M., Kovacs, N.A., Bernier, C.R., Lanier, K.A., Fox, G.E., Harvey, S.C., Wartell, R.M. and Hud, N.V. (2015) History of the ribosome and the origin of translation. *Proceedings of the National Academy of Sciences*, **112**, 15396-15401.

113. Dallas, A., Vlassov, A. and Kazakov, S. (2004), *Artificial nucleases*. Springer, pp. 61-88.

114. Varani, G. and McClain, W.H. (2000) The G· U wobble base pair. *EMBO reports*, **1**, 18-23.

115. Tanaka, Y., Kojima, C., Yamazaki, T., Kodama, T.S., Yasuno, K., Miyashita, S., Ono, A., Ono, A., Kainosho, M. and Kyogoku, Y. (2000) Solution structure of an RNA duplex including a C− U base pair. *Biochemistry*, **39**, 7074-7080.

116. Tanaka, Y., Kojima, C., Yamazaki, T., Kyogoku, Y., Miyashita, S., Ono, A. and Kainosho, M. (1997), *Nucleic acids symposium series*, pp. 271-272.

117. Auffinger, P. and Westhof, E. (1998) Hydration of RNA base pairs. *Journal of Biomolecular Structure and Dynamics*, **16**, 693-707.

118. Fujimoto, M., Fujiyama, K., Kuninaka, A. and Yoshino, H. (1974) Mode of action of nuclease P1 on nucleic acids and its specificity for synthetic phosphodiesters. *Agricultural and Biological Chemistry*, **38**, 2141-2147.

119. Sawai, H., Totsuka, S., Yamamoto, K. and Ozaki, H. (1998) Non-enzymatic, template-directed ligation of 2′–5′ oligoribonucleotides. Joining of a template and a ligator strand. *Nucleic acids research*, **26**, 2995-3000.

120. Sawai, H., Wada, M., Kouda, T. and Nakamura Ozaki, A. (2006) Nonenzymatic Ligation of Short-Chained 2′–5′-or 3′–5′-Linked Oligoribonucleotides on 2′–5′-or 3′–5′-Linked Complementary Templates. *ChemBioChem*, **7**, 605-611.

121. Ciesiołka, J., Michałowski, D., Wrzesinski, J., Krajewski, J. and Krzyżosiak, W.J. (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *Journal of molecular biology*, **275**, 211-220.

122. Kuznetsova, I.L., Zenkova, M.A., Gross, H.J. and Vlassov, V.V. (2005) Enhanced RNA cleavage within bulge-loops by an artificial ribonuclease. *Nucleic acids research*, **33**, 1201-1212.

123. Kaukinen, U., Lyytikäinen, S., Mikkola, S. and Lönnberg, H. (2002) The reactivity of phosphodiester bonds within linear single-stranded oligoribonucleotides is strongly dependent on the base sequence. *Nucleic acids research*, **30**, 468-474.

124. Beloglazova, N., Mironova, N., Konevets, D., Petyuk, V., Sil'nikov, V., Vlasov, V. and Zenkova, M. (2002) Kinetic parameters of cleavage of CpA and UpA sequences in an oligoribonucleotide by compounds functionally mimicking ribonuclease A. *Molecular Biology*, **36**, 869-873.

125. Obianyor, C., Newnam, G., Clifton, B.E., Grover, M.A. and Hud, N.V. (2020) Towards Efficient Nonenzymatic DNA Ligation: Comparing Key Parameters for Maximizing Ligation Rates and Yields with Carbodiimide Activation. *ChemBioChem*, **21**, 3359-3370.

126. Burcar, B., Pasek, M., Gull, M., Cafferty, B.J., Velasco, F., Hud, N.V. and Menor-Salvan, C. (2016) Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angew Chem Int Ed Engl*, **55**, 13249-13253.

127. Cafferty, B.J., Fialho, D.M. and Hud, N.V. (2018) Searching for Possible Ancestors of RNA: The Self-Assembly Hypothesis for the Origin of Proto-RNA. *Nucleic Acids Mol Bi*, **35**, 143-174.

128. Cafferty, B.J. and Hud, N.V. (2014) Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology. *Curr Opin Chem Biol*, **22**, 146-157.

129. Hud, N.V., Cafferty, B.J., Krishnamurthy, R. and Williams, L.D. (2013) The origin of RNA and "my grandfather's axe". *Chem Biol*, **20**, 466-474.

130. El-Sagheer, A.H. and Brown, T. (2010) New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *P Natl Acad Sci USA*, **107**, 15329-15334.

131. Hayden, E.J. and Lehman, N. (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol*, **13**, 909-918.

132. James, K.D., Boles, A.R., Henckel, D. and Ellington, A.D. (1998) The fidelity of template-directed oligonucleotide ligation and its relevance to DNA computation. *Nucleic Acids Res*, **26**, 5203-5211.

133. Gallego, I., Grover, M.A. and Hud, N.V. (2015) Folding and imaging of DNA nanostructures in anhydrous and hydrated deep-eutectic solvents. *Angew Chem Int Ed Engl*, **54**, 6765-6769.

134. Lannan, F.M., Mamajanov, I. and Hud, N.V. (2012) Human telomere sequence DNA in water-free and high-viscosity solvents: G-quadruplex folding governed by Kramers rate theory. *J Am Chem Soc*, **134**, 15324-15330.

135. Li, C., Cafferty, B.J., Karunakaran, S.C., Schuster, G.B. and Hud, N.V. (2016) Formation of supramolecular assemblies and liquid crystals by purine nucleobases and cyanuric acid in water: Implications for the possible origins of RNA. *Physical Chemistry Chemical Physics*, **18**, 20091-20096.

136. Hud, N.V., Fialho, D.M., Karunakaran, S., Greeson, K.W., Martinez, I., Schuster, G.B. and Krishnamurthy, R. (2020) Depsipeptide nucleic acids: prebiotic formation, oligomerization, and self-assembly of a new candidate proto-nucleic acid. *bioRxiv*.

137. Wolfrum, M., Schwarz, R.J., Schwarz, M., Kramer, M. and Richert, C. (2019) Stabilizing DNA nanostructures through reversible disulfide crosslinking. *Nanoscale*, **11**, 14921-14928.

138. El-Sagheer, A.H. and Brown, T. (2010) Click chemistry with DNA. *Chemical Society Reviews*, **39**, 1388-1405.

139. Abe, H., Murayama, D., Kayamori, F. and Inouye, M. (2008) Saccharide-Linked Ethynylpyridine Oligomers: Primary Structures Encode Chiral Helices. *Macromolecules*, **41**, 6903-6909.